ELSEVIER

Biochimica et Biophysica Acta 1443 (1998) 193–197

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Short sequence-paper

A nonsymbiotic hemoglobin gene is expressed during somatic embryogenesis in *Cichorium*[1]

Theo Hendriks [a], Isabelle Scheer [b], Marie-Christine Quillet [b], Béatrice Randoux [b], Bruno Delbreil [b], Jacques Vasseur [b], Jean-Louis Hilbert [b,*]

[a] *Laboratory of Plant Breeding, Agricultural University Wageningen, P.O. Box 386, 6700 AJ Wageningen, The Netherlands*
[b] *Laboratoire de Physiologie Cellulaire et Morphogenèse Végétales, USTL/INRA, Université des Sciences et Technologies de Lille, Bâtiment SN2, 3ème étage, F-59655 Villeneuve d'Ascq Cedex, France*

Received 30 July 1998; received in revised form 17 September 1998; accepted 24 September 1998

Abstract

After differential screening of a cDNA library corresponding to genes expressed during the early stages of somatic embryogenesis in leaf tissue from the *Cichorium* hybrid '474' (*C. intybus* L., var. *sativum*×*C. endivia* L., var. *latifolia*) a nonsymbiotic hemoglobin cDNA was obtained. Studies of the expression of the gene corresponding to this clone by Northern blot analysis suggest that in *Cichorium* a nonsymbiotic hemoglobin gene is specifically expressed under somatic embryogenesis-inducing conditions, and that its expression is not related to stress caused by wounding or tissue culture conditions.

Keywords: Differential screening; Nonsymbiotic hemoglobin; Somatic embryogenesis; (*Cichorium*)

In higher plants two families of hemoglobins are distinguished; the symbiotic and nonsymbiotic hemoglobins [1,2]. Symbiotic hemoglobins, or leghemoglobins, are only detected in nitrogen-fixing nodules formed upon the interaction between bacteria and its host plants, whereas nonsymbiotic hemoglobins are more widely distributed and are present in both monocot and dicot plants. In nitrogen-fixing nodules the function of symbiotic hemoglobins is to transport oxygen to the bacteroids. At the low oxygen concentration in infected cells, to avoid damage to the oxygen-sensitive nitrogenase, leghemoglobins ensure an adequate supply of oxygen for bacterial respiration [3]. The function of nonsymbiotic hemoglobins is not known, but since their encoding genes are specifically expressed in tissue were the oxygen supply is limited or that are metabolically very active [1,4–7], they too may act as oxygen carriers.

Here we report the identification of a nonsymbiotic hemoglobin cDNA obtained after differential screening of a cDNA library corresponding to genes expressed during the induction of somatic embryogenesis in leaf tissue from the *Cichorium* hybrid '474' (*C. intybus* L., var. *sativum*×*C. endivia* L., var. *latifolia*).

To induce somatic embryogenesis, leaf fragments from 6 weeks-old plantlets were cultured in darkness for 5 days at 35°C in 50 ml of an agitated (50 rpm) liquid basal M17 medium [8], supplemented with 60

* Corresponding author. Fax: +33 03 2033-7244; E-mail: hilbert@univ-lille1.fr

[1] The EMBL accession number for the *CHI3206* sequence is AJ007507.

0167-4781/98/$ – see front matter
PII: S0167-4781(98)00206-1

mM sucrose and 330 mM glycerol. The presence of glycerol delayed the first divisions of embryogenic cells and allowed a relative synchronisation of the first divisions after transferring the leaf fragments to the same medium without glycerol [9]. Poly(A^+) RNA (1 μg) isolated from 3 days-old induced leaf tissue was used to construct a cDNA library in phage lambda Uni-ZAP XR from Stratagene (La Jolla, CA). *Escherichia coli* XL-1 blue and SOLR (Stratagene, La Jolla, CA) were used as hosts for lambda Uni-ZAP XR and pBluescript plasmids, respectively. Clones obtained after differential screening of the library with radioactive cDNA probes corresponding to mRNAs isolated from leaf tissue cultured for 3 days in embryo-inducing medium and to mRNAs isolated from non-induced leaf tissue, were subjected to two more rounds of differential screening before further characterisation.

Comparison of the nucleotide sequences of the clones obtained after differential screening with sequences in the EMBL GenBank revealed that one of them – *CHI3206* – is highly similar to plant non-

```
1     GG CAC GAG ACA GAG ATG GGT TTT AGT GAG AAG CAA GAA GCA TTG GTG AAG GAG TCA   56
1                        M   G   F   S   E   K   Q   E   A   L   V   K   E   S    14

57    TGG GAG GTG ATG AAG CAA GAC ATT CCT GCT CTT AGT CTC TAT CTT TAC GCG ATG ATA  113
15    W   E   V   M   K   Q   D   I   P   A   L   S   L   Y   L   Y   A   M   I    33

114   CTG GAG ATA GCC CCG GAA GCA AGA GGG TTG TTT TCA TTT CTG AAA GAC ACA AAT GTA  170
34    L   E   I   A   P   E   A   R   G   L   F   S   F   L   K   D   T   N   V    52

171   ATT CCC CAA AAT AAC CCC AAG CTC AAG TCC CAT GCT GTT AAG GTT TTC AAG ATG GTA  227
53    I   P   Q   N   N   P   K   L   K   S   H   A   V   K   V   F   K   M   V    71

228   TGT GAG TCA GCA ATT CAA CTG AGG GAG AAA GGT GAG GTT GTG GTT TCT GGT TCT ACA  284
72    C   E   S   A   I   Q   L   R   E   K   G   E   V   V   V   S   G   S   T    90

285   CTC AAG TAT TTG GGA TCC GTT CAT CTT GAG AAA GGA ATC GTC GAT CCT CAA TTT GAG  341
91    L   K   Y   L   G   S   V   H   L   E   K   G   I   V   D   P   Q   F   E   109

342   GTG GTG AAA GAG GCT TTA ATA AGA ACA GTG GCG AAG GCA ATG GGA GAG AAA TGT AGT  398
110   V   V   K   E   A   L   I   R   T   V   A   K   A   M   G   E   K   C   S   128

399   GAA GAA ATG AAG AGT GCT TGG TCT GAA GCT TAT GAT GAG TTG GCT GAT GCC ATC AAA  455
129   E   E   M   K   S   A   W   S   E   A   Y   D   E   L   A   D   A   I   K   147

456   ACT GAA ATG AAG AAA GAG GCT GCA GAA ACT CAA ACT CAC ATT TGA ATA CTT GAG AGA  512
148   T   E   M   K   K   E   A   A   E   T   Q   T   H   I                      161

513   AAT TGA TGG TCT TCG AAA TTC GAA TAA AGT GGA AAA CAA CAA TAT ATT GAT ATA TGT  569

570   TTA TTA CAA AAC GGT ATT TTG TTC TTA AAT CAA AAA AAA AAA AAA AA               616
```

Fig. 1. Nucleotide and deduced amino acid sequence of the cDNA clone *CHI3206* from the *Cichorium* hybrid '474'. The putative start and stop codons and the sequence corresponding to eukaryotic polyadenylation consensus sequence are underlined. Both strands of plasmid DNA of clone *CHI3206* were sequenced between the T3 and T7 promoter sequences flanking the pBluescript cloning site. The DNA fragments were fluorescently labelled using the Prism AmpliTaq FS Dye Primer kit (PE Applied Biosystems, Foster City, CA) and the sequence determined using a LI-COR Long Read IR DNA sequencer in combination with the Base Image IR L-4000-05 software (LI-COR, Lincoln, NE).

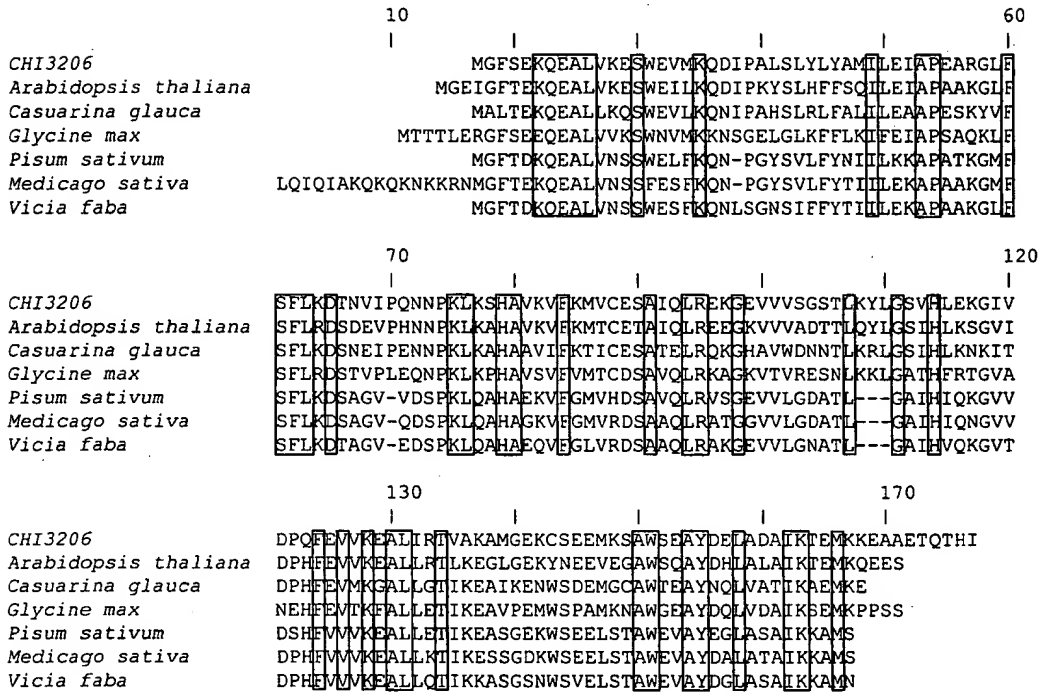

Fig. 2. Alignment of the predicted amino acid sequence encoded by clone *CHI3206* and five hemoglobins from different plant species. Sequences were obtained from the GenBank database accession numbers U94999 (*Arabidopsis thaliana*), X77695 (*Casuarina glauca*), Z54157 (*Vicia faba*), AB01831 (*Pisum sativum*), X54089 (*Medicago sativa*) and U47143 (*Glycine max*). Alignment of sequences was done using CLUSTALW at the Network Protein Sequence @nalysis (IBCP, France).

symbiotic hemoglobins. The nucleotide and deduced amino acid sequences of clone *CHI3206* are shown in Fig. 1. The nucleotide sequence of the 616 bp revealed an open reading frame of 486 bp and 14 bp and 102 bp 5′ and 3′ flanking regions. The *CHI3206* deduced protein is 161 amino acids long, including the N-terminal methionine and has a calculated molecular mass of 18 024 Da and a p*I* of 5.35. The predicted protein contains no signal peptide and no putative N-linked glycosylation sites. Sequences alignment of the deduced protein and hemoglobins from five different plant species in the SwissProt database (Fig. 2), shows 46–63% identity and 71–88% similarity.

Total RNA from non-induced leaf tissue and from leaf tissue explants taken from embryo cultures was subjected to Northern blot analysis using *CHI3206* cDNA as a probe (Fig. 3A, d0 to d5+3). Hybridisation of *CHI3206* probe with a 0.7 kb transcript was observed in leaf tissue upon embryo induction, and also in leaf tissue after transfer to expression medium, albeit at a much lower level. The level of *CHI3206* mRNA increased from day 1 onward to a maximum at day 3 and decreased again thereafter (Fig. 3A, d1 to d5+3). Scanning the blots suggested that the level of *CHI3206* mRNA at day 3 was about 35-fold higher when compared to the level at day 1 (Fig. 3C). At day 4 and day 5 the levels were still about 20- and 10-fold higher respectively, whereas 3 days after transfer to the expression medium (d5+3) the level was about equal to that at day 1. No hybridisation of *CHI3206* was observed with RNA isolated from non-induced leaf tissue (Fig. 3A, d0), nor with RNA from leaf tissue of a non-embryogenic cultivar *Orchies* of *C. intybus* when cultured under embryogenesis-inducing or -expression conditions (Fig. 3A, d0 to d5+3). These results suggest that in *Cichorium* '474' the nonsymbiotic hemoglobin gene corresponding to *CHI3206* is specifically expressed under somatic embryogenesis-inducing conditions, and that its expression is not related to stress caused by wounding or tissue culture conditions. The high level of transcripts present at day 3 suggests that the expression of the hemoglobin gene

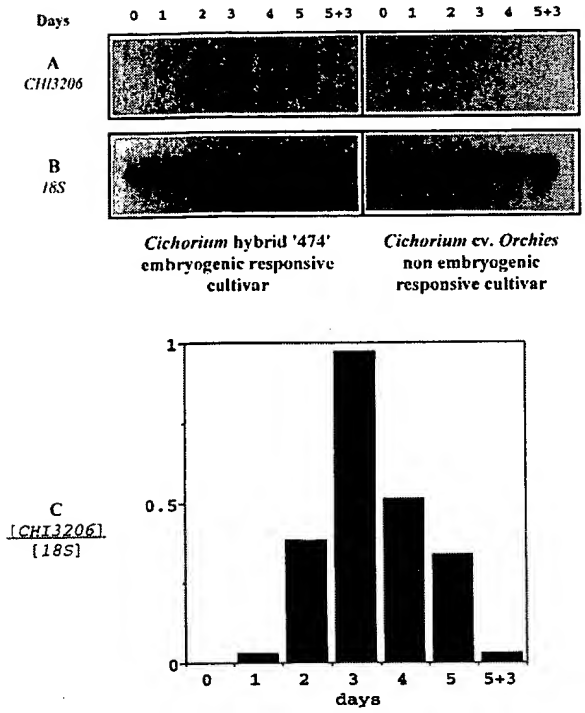

Fig. 3. Northern blot analysis of the *CHI3206* hemoglobin gene expression during somatic embryogenesis in *Cichorium* hybrid '474'. (A and B_ Autoradiograms of Northern blots hybridised with ^{32}P probes corresponding respectively to *CHI3206* (A) and a 18S genomic fragment *Xba*I-*Eco*RI (1.4 kb) from *Cichorium intybus* (B). Total RNA (10 μg) was isolated from non-induced leaf tissue (d0, lanes 1, 8), leaf tissue subjected for 1, 2, 3, 4, or 5 days to somatic embryogenesis induction medium (lanes 2–6 and 9–12), and leaf tissue subjected for 5 days to somatic embryogenesis induction and cultured for 3 days in expression medium (lanes 7, 13) of the *Cichorium* hybrid '474' embryogenic responsive line (lanes 1–7) or of the *Cichorium* cv. *Orchies* non-embryogenic responsive cultivar (lanes 8–13). Total RNA was separated by agarose electrophoresis and transferred to GeneScreen nylon membranes (NEN Biomedical, Boston, MA). The probes for *CHI3206* and 18S were ^{32}P labelled by random priming with the T7 QuickPrime Kit (Pharmacia Biotech, Freiburg, Germany). (C) Transcription level of *CHI3206* during induction of somatic embryogenesis in *Cichorium* hybrid '474'. The lanes of the autoradiograms in (A) and (B) were scanned using a Sharp JX330 scanner connected to a Sun Sparc 20 microcomputer. The TIFF files were analysed using the Whole Band Analyzer software from BioImage (BioImage Corporation, Ann Arbor, MI). Small variations between total mRNA samples were corrected by using the ratio of hybridisation signals of *CHI3206* hemoglobin and 18S rRNA hybrid.

is correlated with the initiation of somatic embryogenesis rather then with somatic embryogenesis per se. The presence of the hemoglobin mRNA after transfer of the leaf tissue explants to the expression medium may be due to asynchronous initiation of somatic embryogenesis in the expression medium [8].

Upon the first discovery of nonsymbiotic hemoglobins [10], Appleby et al. [11] suggested that nonsymbiotic hemoglobins may be involved in facilitating the diffusion of oxygen and/or the sensing the levels of oxygen available in roots. They hypothesised that under normal conditions hemoglobin would be oxygenated and that an increased level of deoxyferous hemoglobin would trigger a response to hypoxic conditions. In barley and maize, the expression of a nonsymbiotic hemoglobin gene was induced under hypoxic conditions [6,10], and in *Casuarina glauca* too the expression seemed restricted to internal tissues, where oxygen supply may be limiting [4]. During somatic embryogenesis the leaf explant tissue is submerged in liquid medium, and though the cultures are agitated, hypoxic conditions may result, particular at the elevated temperature (35°C) conditions. These hypoxic conditions then may be responsible for inducing the expression of the hemoglobin gene in the embryogenic responsive cultivar. However, in leaf tissue explants from a non-embryogenic *C. intybus* cv. *Orchies* cultured under similar conditions the hemoglobin gene is not expressed (Fig. 3A). The induction of embryogenesis in *Cichorium* '474' is accompanied by the reactivation of somatic cells [8], a developmental switch that undoubtedly requires a complex array of energy-demanding processes, and it may be that the hemoglobin is involved in the energy generation. The question raised is whether the expression of the hemoglobin gene is triggered by the local oxygen limiting conditions in the reactivated cells or whether the induction of the hemoglobin gene expression requires a more specific signal. Promoters of symbiotic and nonsymbiotic hemoglobin genes were shown to direct both root-specific and nodule-specific expression of the β-glucuronidase gene in transgenic plants, indicating that these promoters share conserved *cis*-acting DNA motifs involved in the regulation of gene expression [4,12]. Initially, symbiotic hemoglobins were thought to be typical late nodulins, being present only after formation of the nodule [3]. However, it was found recently in *Vicia sativa* (vetch) that the expression of the leghemoglobin gene *VsLB1* may actually be one of the first responses upon *Rhizobium*-legume interaction

[5]. Within 1 h after the application of *Rhizobium* lipo-chitin oligosaccharides, called Nod factors [13], *VsLB1* mRNA was detectable in the root hairs of the susceptible zone. It was suggested that the leghemoglobin may facilitate oxygen supply to support mitochondrial respiration in the rapidly respiring root hairs [5]. It may be that in *Cichorium* both somatic embryogenesis and hemoglobin gene expression are induced in response to the same signal molecule(s), similar to the induction of root hair deformation and the expression of the leghemoglobin gene by Nod factors in the water-logged root hairs in the *Rhizobium–Vicia* interaction [5]. In this respect, the possible involvement of Nod factor-like signal molecules in somatic embryogenesis [14] is particularly interesting.

Somatic embryogenesis is the result of dedifferentiation and reacquisition of a morphogenic competence. The transient expression of a nonsymbiotic hemoglobin gene during the induction of somatic embryogenesis may be related either to the dedifferentiation process or to the reacquisition of embryogenic competence. In a previous study [15], three 18.1 kDa acidic proteins, $pI \approx 5$, were found to be transiently present at a low level in leaf tissue explants during the induction phase of *Cichorium* '474' somatic embryogenesis. The physico-chemical parameters of these proteins are similar to the putative hemoglobin encoded by *CHI3206*, and we are currently in pursuit of their identity. Further analysis of the expression of the *CHI3206* gene, in particular by in situ hybridisation, may facilitate the unravelling of the role of hemoglobin during *Cichorium* somatic embryogenesis.

This work was in part supported by a 'Contrat Plan-Etat-Region' to the Laboratoire de Physiologie Cellulaire et Morphogenèse Végétales and by a post-doctoral fellowship of the Conseil Régional Nord-Pas de Calais to Theo Hendriks. We thank Dr. Annick Bellamy for providing the chicory 18S DNA.

References

[1] C.R. Andersson, E. Østergaard, D.J. Llewellyn, E.S. Dennis, W.J. Peacock, Proc. Natl. Acad. Sci. USA 93 (1996) 5682–5687.
[2] C.A. Appleby, Sci. Progr. 76 (1992) 365–398.
[3] C.A. Appleby, Plant Physiol. 35 (1984) 443–478.
[4] K. Jacobsen-Lyon, E. Østergaard Jensen, J.-E. Jørgensen, K.A. Macker, W.J. Peacock, E.S. Dennis, Plant Cell 7 (1995) 213–223.
[5] R. Heidstra, G. Nilsen, F. Martinez-Abarca, A. Van Kammen, T. Bisseling, Mol. Plant-Microbe Interact. 10 (1997) 215–220.
[6] E.R. Taylor, X.Z. Nie, A.W. MacGregor, R.D. Hill, Plant Mol. Biol. 24 (1994) 853–862.
[7] C.D. Drew, Annu. Rev. Plant Physiol. Plant Mol. Biol. 48 (1997) 223–250.
[8] T. Dubois, M. Guedira, J. Dubois, J. Vasseur, Protoplasma 162 (1991) 120–127.
[9] A.S. Robatche-Claive, J.-P. Couillerot, J. Dubois, T. Dubois, J. Vasseur, C.R. Acad. Sci. Paris 314 (1992) 371–377.
[10] D. Bogusz, C.A. Appleby, J. Landsmann, E.S. Dennis, M.J. Trinick, J.W. Peacock, Nature 331 (1988) 178–180.
[11] C.A. Appleby, D. Bogusz, E.S. Dennis, W.J. Peacock, Plant Cell Env. 11 (1988) 359–367.
[12] D. Bogusz, D.J. Llewellyn, S. Craig, E.S. Dennis, C.A. Appleby, W.J. Peacock, Plant Cell 2 (1990) 633–641.
[13] R.W. Carlson, N.P.J. Price, G. Stacey, Mol. Plant-Microbe Interact. 7 (1995) 684–695.
[14] A.M. De Jong, R. Heidstra, H.P. Spaink, M.V. Hartog, E.A. Meijer, T. Hendriks, F. LoSchiavo, M. Terzi, T. Bisseling, A. Van Kammen, S.C. de Vries, Plant Cell 5 (1993) 615–620.
[15] C. Boyer, J.-L. Hilbert, J. Vasseur, Plant Sci. 93 (1993) 41–53.

Plant Cell Physiol. 43(11): 1351–1358 (2002)

Expression of Symbiotic and Nonsymbiotic Globin Genes Responding to Microsymbionts on *Lotus japonicus*

Toshiki Uchiumi [1,5], Yoshikazu Shimoda [2], Tomoko Tsuruta [2], Yumi Mukoyoshi [1], Akihiro Suzuki [1], Keishi Senoo [3], Shusei Sato [4], Tomohiko Kato [4], Satoshi Tabata [4], Shiro Higashi [1] and Mikiko Abe [1]

[1] *Department of Chemistry and BioScience, Kagoshima University, Kagoshima, 890-0065 Japan*
[2] *Graduate School of Science and Technology, Kagoshima University, Kagoshima, 890-0065 Japan*
[3] *Graduate School of Agricultural and Life Sciences, University of Tokyo, Tokyo, 113-8657 Japan*
[4] *Kazusa DNA Research Institute, Chiba, 292-0812 Japan*

Leguminous plants have both symbiotic and nonsymbiotic hemoglobin (sym- and nonsym-Hb) genes. Three symbiotic (*LjLb1, 2, 3*) and one nonsymbiotic (*LjNSG1*) Hb genes were isolated from a genomic library of *Lotus japonicus* MG20 Miyakojima. Two motif sequences (AAAGAT and CTCTT) critical for nodule specific expression were conserved on the 5′-upstream sequence of *LjLb1, 2* and *3*. The 5′-upstream region of *LjNSG1* contained the sequence consensus for nonsym-Hb. RT-PCR with specific primer sets for each *LjLb* gene showed that all the sym-Hb genes (*LjLb1, 2, 3*) were expressed specifically and strongly in root nodules. The expression of *LjLb1, 2* and *3* could not be detected in root, leaf or stem of a mature plant, whereas low level expression was detected in seedlings by RT-PCR. This suggests that sym-Hbs may have another unknown function besides being oxygen transporter for the microsymbiont in symbiotic nitrogen fixation. The expression of *LjNSG1*, examined with RT-PCR, was detected at low level in root, leaf and stem. The expression of *LjNSG1* was enhanced in root nodules, whereas it was repressed in roots colonized by mycorrhizal fungi *Glomus* sp. R10. The repression of the nonsym-Hb gene was also observed in the roots of *Medicago sativa* colonized by *Glomus* sp. R10.

Keywords: Expression — Hemoglobin — *Lotus japonicus* — Mycorrhiza — Nodule.

Abbreviations: AM, arbuscular mycorrhiza; Hb, hemoglobin; Lb, leghemoglobin; nonsym-Hb, nonsymbiotic hemoglobin; sym-Hb, symbiotic hemoglobin.

The nucleotide sequences reported in this paper have been submitted to DDBJ under accession numbers AP004624, AP004625, AP004626, AP004627, AP004628 and AP004629.

Introduction

Hemoglobin (Hb) in legumes, so called leghemoglobin (Lb), exists abundantly in root nodules and regulates oxygen concentration to create a micro-aerobic environment for nitrogenase activity of rhizobia in the symbiotic zone. Hbs have also been found in the root nodules of nitrogen-fixing nonlegumes such as *Casuarina glauca*, *Myrica gale* and *Alnus glutinosa*. These Hbs, which are nodule specific or associate with a symbiotic interaction with microorganisms, are called "symbiotic hemoglobins (sym-Hbs)". Lbs, i.e. sym-Hbs in legumes, are encoded by a family of genes and multiple isomers exist in nodules.

Other Hbs showing only a limited amino acid sequence similarity to the sym-Hbs have been identified in non-nodulating plant *Hordeum*, *Triticum*, *Arabidopsis*, *Zea* and *Oryza* (Arredondo-Peter et al. 1998). These findings suggest that Hbs are contained in all plants, and the Hbs in non-nodule tissues are called "nonsymbiotic hemoglobins (nonsym-Hbs)". Nonsym-Hbs have also been identified in legumes, *Glycine max* (Andersson et al. 1996) and *Medicago sativa* (Seregélyes et al. 2000) and in nodulating non-legumes *Casuarina glauca* (Jacobsen-Lyon et al. 1995). A nonsym-Hb gene in *M. sativa*, *Mhb1*, is expressed under hypoxia as are other nonsym-Hbs (Seregélyes et al. 2000). However, it is not clear whether nonsym-Hb has any role in symbiosis with microorganisms.

In this report, genomic clones of three symbiotic and one nonsymbiotic Hb genes were isolated from the genomic library of a model legume *Lotus japonicus* MG20 Miyakojima (Sato et al. 2001). To examine the response of these Hb genes to microsymbionts, the expression of each sym- and nonsym-Hb gene was analysed by RT-PCR. We focused on the expression of Hb genes in plants that are in symbiosis with *Mesorhizobium loti* and *Glomus* sp. The expression of the nonsym-Hb gene of *M. sativa* was also examined in plants that are in symbiosis with *Glomus* sp. for comparison with that of *L. japonicus*. This is the first report about the expression pattern of nonsym-Hb gene within the plant that is in symbiosis with mycorrhizal fungi.

Results

Structure of sym-Hb genes of L. japonicus

Two sets of PCR primer pairs, LjLbF1/LjLbR1 and LjN77F1/LjN77R1, were used for amplification of the part of genomic sym-Hb genes of *L. japonicus*. Three DNA fragments were amplified by PCR using genomic DNA of *L. japonicus*

[5] Corresponding author: E-mail, uttan@sci.kagoshima-u.ac.jp; Fax, +81-99-285-8163.

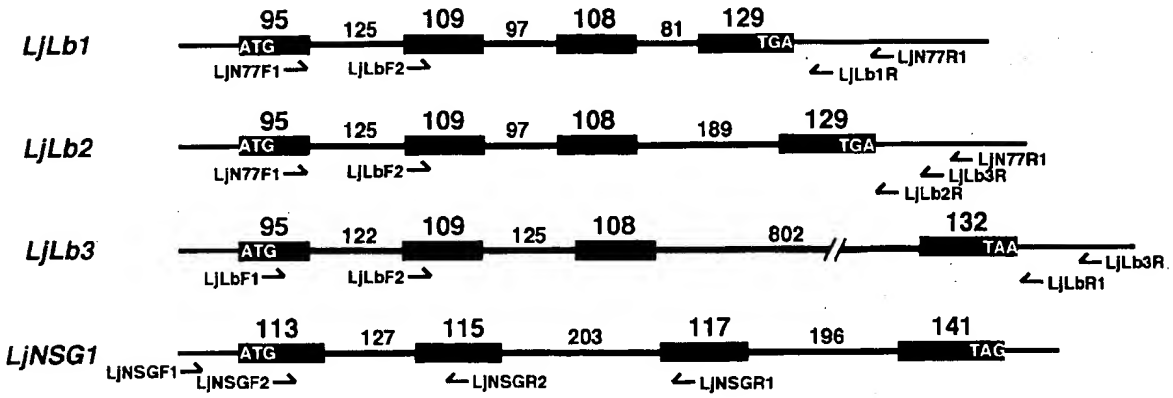

Fig. 1 Intron and exon structure of sym- and nonsym-Hb genes of *L. japonicus*. Exons are indicated by black boxes. The numbers indicate the size of exons and introns in bp. The small arrows indicate the annealing positions of primers for PCR.

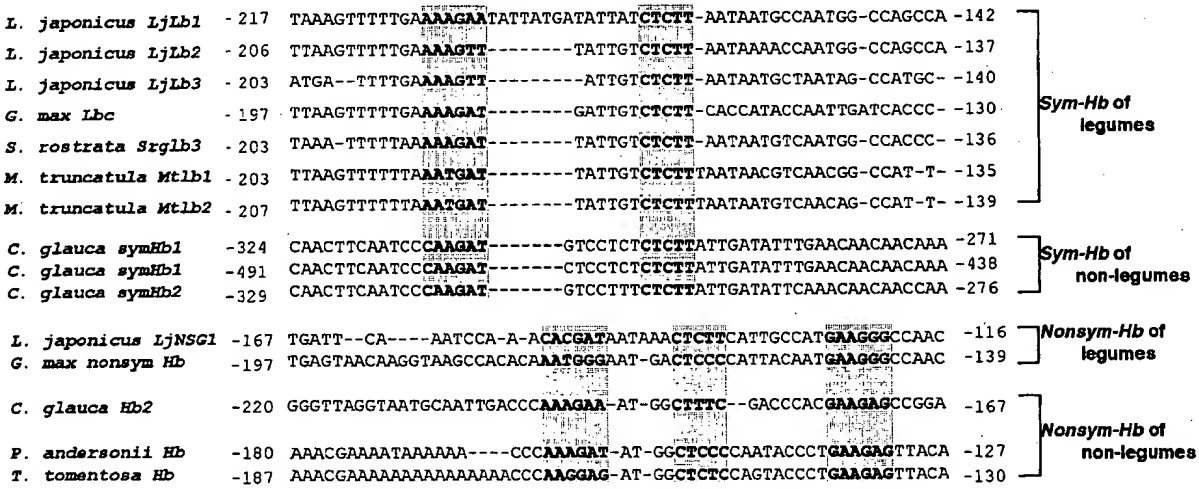

Fig. 2 Sequence alignment of the promoter regions of sym-Hb and nonsym-Hb genes. All the sequences are numbered from the translation initiation codon ATG. The motifs resembling nodulin motif (AAAGAT and CTCTT) were found in 5′-upstream of the three sym-Hb genes (*LjLb1, LjLb2, LjLb3*) of *L. japonicus*. The motif GAGGG was conserved in 5′-upstream of the nonsym-Hb genes of plants. The motifs are indicated by boldface in the gray boxes.

Miyakojima MG20 as a template, suggesting that at least three independent sym-Hb genes exist on the genome. Another three sets of primer pairs, LjLbF2/LjLb1R, LjLbF2/LjLb2R and LjLbF2/LjLb3R, were redesigned according to the sequences of the three fragments generated by PCR using primer pairs of LjLbF1/LjLbR1 and LjN77F1/LjN77R1, and were used for screening the 3D-pool of the genomic library constructed by the TAC vector (Liu et al. 1999, Sato et al. 2001). As a result of screening, three independent sym-Hb genes were identified and were referred to as *LjLb1*, *LjLb2* and *LjLb3* (Fig. 1). The TAC clones on which *LjLb1* and *LjLb2* were located adjoined each other. Furthermore, *LjLb3* was mapped approx. 3 cM away from *LjLb1* and *LjLb2* on the same linkage group (linkage group 5, marker names are TM0089 for *LjLb1* and *LjLb2* and TM0090 for *LjLb3*; http://www.kazusa.or.jp/lotus/). The genomic sequences of the three sym-Hb genes were decided including a 5′-promoter region and 3′-UTR. Although the sequences of the four exons and two of the three introns were conserved, the third intron was specific to each sym-Hb gene. No significant homology could be detected in the 3′-UTR except for the putative signal sequence (AATAAA) for poly-A addition. The sequences of the 5′-upstream regions showed high homology with the promoter sequence of sym-Hb gene of other leguminous plants. The nodulin motif (AAGAT and CTCTT) like sequences were found in this region (Fig. 2). For *LjLb1*, the spacing (14 bp) between the motif sequences differs from that in other sym-Hb genes *LjLb2* (6 bp) and *LjLb3* (5 bp).

Structure of nonsym-Hb gene of L. japonicus

Two sets of primer pairs for amplification of nonsym-Hb gene, LjNSGF1/LjNSGR1 and LjNSGF2/LjNSGR2, were

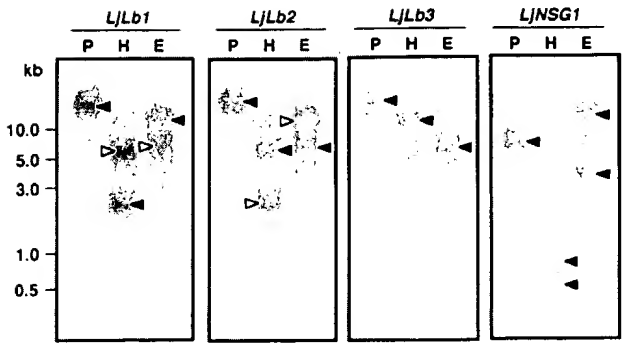

Fig. 3 Southern hybridization probed with the partial fragments of the each globin gene. Genomic DNA of *L. japonicus* was digested by *Pst*I (lane P), *Hin*dIII (lane H) and *Eco*RI (lane E), respectively. The restriction fragments hybridized with the respective probes are indicated by closed arrowheads and cross hybridized fragments are indicated by open arrowheads.

designed according to the sequence of the nonsym-Hb homologue appearing in seedling expressed sequence tags (ESTs) of *L. japonicus* (Asamizu et al. 2000). Two DNA fragments were amplified by PCR with each primer pair and the sequences of the amplified products supported the fact that these DNA fragments were derived from the same gene on the genome. From the 3-D pool of the genomic library, one TAC clone was screened using the primer pair LjNSGF1/LjNSGR1 for nonsym-Hb gene. The Hb gene on this clone showed high homology with known nonsym-Hb genes and was referred to as *LjNSG1* (Fig. 1). *LjNSG1* was mapped on the linkage group 3 (marker name is TM0091, http://www.kazusa.or.jp/lotus/). The highly conserved sequence (GAAGGG) among the nonsym-Hb genes existed on the 5'-upstream region of *LjNSG1* and one of the nodulin motif sequences (CTCTT) was also found (Fig. 2).

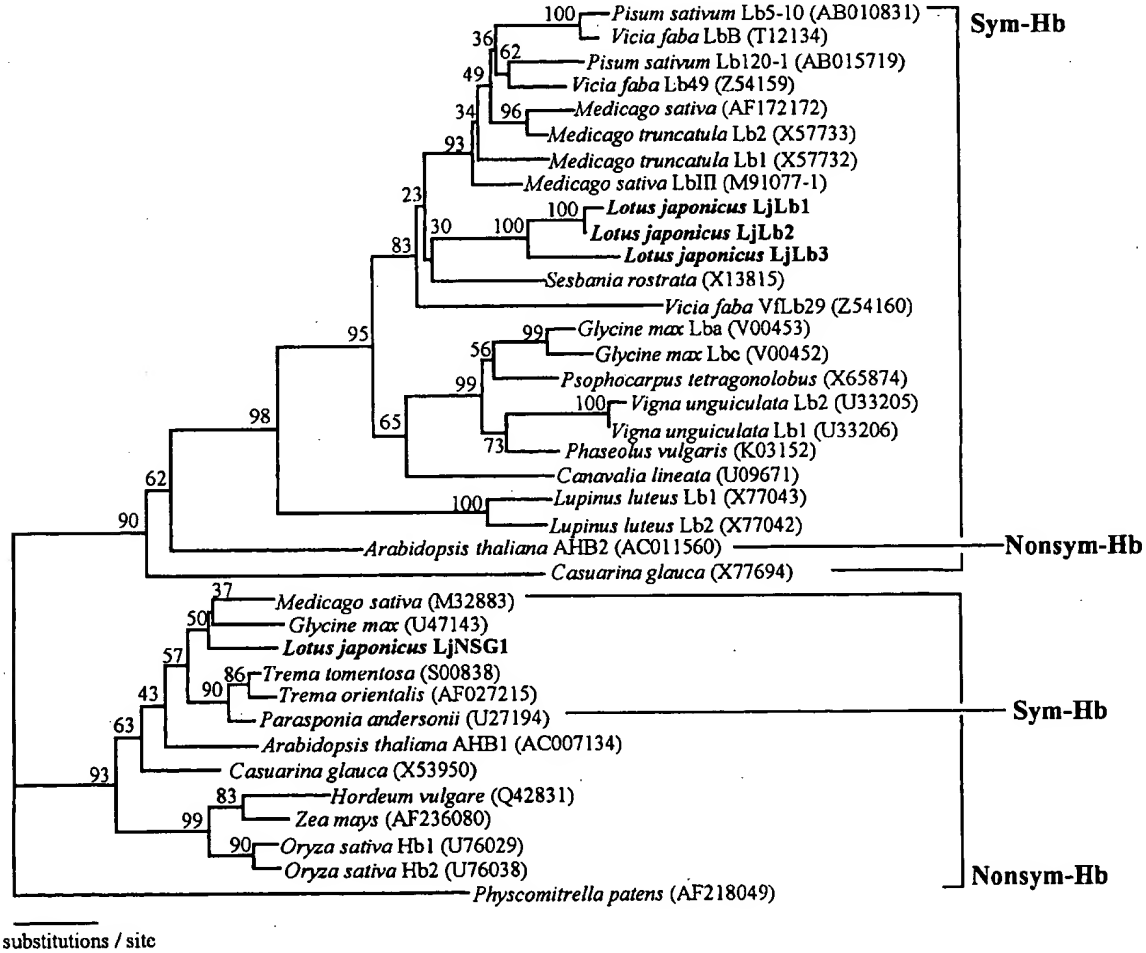

Fig. 4 Phylogenetic tree based on the amino acid sequences of plant hemoglobins. The tree was constructed by neighbour-joining method. The numbers on the branches show bootstrap probabilities (as percentages) determined from 1,000 resamplings. The database accession numbers are indicated in parentheses after the plant names.

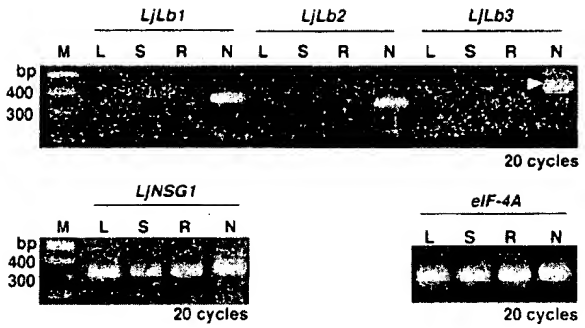

Fig. 5 Expression of sym-Hb and nonsym-Hb genes in different tissues of a mature plant of *L. japonicus*. Total RNAs were isolated from leaf (lane L), stem (lane S), root (lane R), and nodule (lane N). RT-PCR was performed using specific primers for each sym- and nonsym-Hb genes. The PCR product derived from transcript of *LjLb3* is indicated by an arrowhead. The transcripts of *LjLbs* were detected only in nodules, whereas the transcript of *LjNSG1* was detected in all the tissues tested. The *eIF-4A* of *L. japonicus* was used as a positive control. The number of amplification cycles was 20. Lane M, size marker.

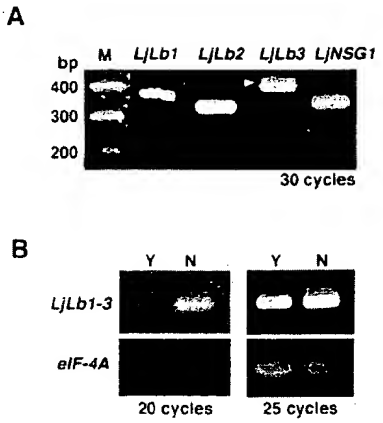

Fig. 6 Expression of sym-Hb genes in young shoots and nodules of *L. japonicus*. (A) Total RNA was prepared from young shoots of three days after germination. RT-PCR was performed using specific primers for each Hb gene. The PCR product derived from the transcript of *LjLb3* is indicated by an arrowhead. Lane M, size marker. (B) Total RNA was prepared from young shoots (Y) or mature nodules (N). The degenerate primer pair for the three *LjLbs* was used for amplification. The *eIF-4A* was used as a positive control.

Southern hybridization probed with sym- and nonsym-Hb genes of L. japonicus

Southern hybridization probed with sym- and nonsym-Hb genes was performed against genomic DNA of *L. japonicus* (Fig. 3). Total genomic DNA was digested with restriction endonucleases *Pst*I, *Hin*dIII and *Eco*RI, and separated on 0.7% agarose gel. After blotting onto the nylon filter, hybridizations were carried out using ^{32}P-labeled probes prepared by PCR using the corresponding TAC clone as a template. The hybridization profiles obtained from *LjLb1*- and *LjLb2*-probes were identical each other, because *LjLb1* and *LjLb2* were located on the same *Pst*I fragment, and cross hybridization occurred for *Hin*dIII and *Eco*RI fragments (Fig. 3, *LjLb1*, *LjLb2*). Hybridization profiles obtained from *LjLb3*- and *LjNSG1*-probe were specific to each probe (Fig. 3, *LjLb3*, *LjNSG1*). All the hybridized fragments detected were consistent with the restriction maps of each TAC clone carrying sym- and nonsym-Hb genes.

Phylogeny of Sym- and Nonsym-Hbs of L. japonicus with Hbs in other plants

A phylogenetic tree was constructed comparing the amino acid sequences with a known plant hemoglobin, using the neighbour-joining method in the ClustalW program (Fig. 4). The predicted amino acid sequences of the three sym-Hbs (LjLb1, LjLb2 and LjLb3) of *L. japonicus* form a cluster within a large cluster of sym-Hbs and with VfLb29 of the gene expression induced by the infection of the mycorrhizal fungi (Frühling et al. 1997). The predicted amino acid sequence of LjNSG1, nonsym-Hb of *L. japonicus*, belonged to a cluster with other nonsym-Hbs of leguminous plants.

Expression of sym- and nonsym-Hb genes in nodules and in other tissues

The expression of sym- and nonsym-Hb genes in nodules and in other tissues of *L. japonicus* was analysed by RT-PCR. For detection of the transcripts of *LjLb1*, *LjLb2*, *LjLb3* and *LjNSG1*, primer pairs of LjLbF2/LjLb1R, LjLbF2/LjLb2R, LjLbF2/LjLb3R and LjNSGF1/LjNSGR1 were used, respectively. The primer pair of LjLbF2/LjLb3R generated two cDNA fragments; one was a derivative from a transcript of *LjLb2* and the other was from *LjLb3* (Fig. 5, *LjLb3* lane N; Fig. 6A, lane *LjLb3*). Although the primer pair of LjLbF2/LjLb3R is not specific to *LjLb3* transcript, the pair is useful for detection of the *LjLb3* transcript. RT-PCR using primer pairs for sym-Hb genes (*LjLb1*, *2* and *3*) showed that all the sym-Hb genes were expressed specifically and strongly in root nodules (Fig. 5). The significant difference on gene expression did not exist in the case of the three sym-Hb genes of *L. japonicus*. The expression of the sym-Hb genes could not be detected in root, leaf and stem of a mature plant even though the number of amplification cycles was increased to 35. However, the low level expression was detected in seedlings 3 d after germination (Fig. 6). The transcripts from all the sym-Hb genes were detected (Fig. 6A) and the expression level in young shoots was estimated to be very low compared with that in nodules (Fig. 6B). The expression of *LjN16*, one of the representative late nodulin genes (Kapranov et al. 1997), could not be detected, indicating that there was no RNA contamination from nodules into the RNA prepared from young seedlings (data not shown). This suggests that sym-Hb has unknown function in young tissues other than oxygen transporter for the microsymbiont in nodules.

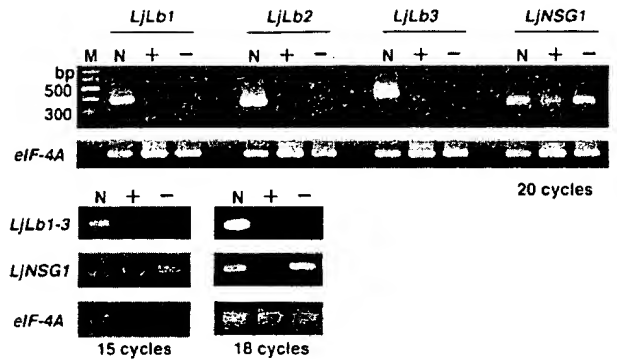

Fig. 7 Expression of sym-Hb and nonsym-Hb genes in nodules and mycorrhiza. Total RNAs were isolated from nodules (lane N), roots colonized by *Glomus* sp R-10 (lane +) and roots not infected with *M. loti* and *Glomus* (lane −). The transcripts of *LjLbs* were detected only in nodules. The amount of the transcript of *LjNSG1* seems to be low in mycorrhiza (*LjNSG1*, lane +) compared with that in the root without infection (*LjNSG1*, lane −). *Ljlb1-3* indicates the amplified product using the degenerate primer pair for the three *LjLbs*. The *eIF-4A* of *L. japonicus* was used as a positive control.

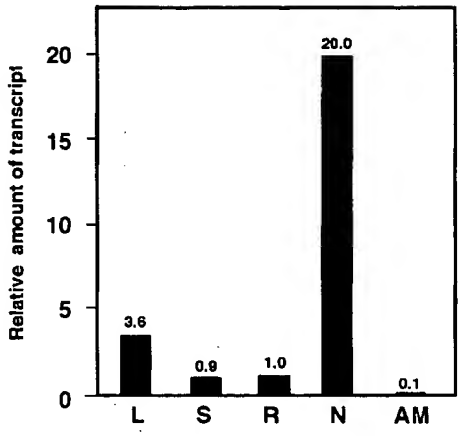

Fig. 8 Relative amounts of the transcript of *LjNSG1*. The amounts of the transcript in different tissues were estimated by quantitative real time RT-PCR and indicated as the amount in the normal root was 1.0. The amount of the transcript of *LjeIF-4A* was used for normalization. The values indicate the average of three independent experiments. The expression of *LjNSG1* in root nodules (N) was enhanced and that in mycorrhiza (AM) was repressed compared with normal roots (R). L, leaf; S, stem; R, normal root (root without *Glomus* sp. R-10); N, nodules; AM, root colonized by *Glomus* sp. R-10.

The expression of nonsym-Hb gene (*LjNSG1*) was detected at low level in root, leaf and stem by RT-PCR (Fig. 5). *LjNSG1* was also expressed in nodules and the level of expression was enhanced compared with that in the other tissues (Fig. 5). The amount of *LjNSG1* transcripts in total RNA was estimated by quantitative RT-PCR and compared between nodules and other tissues. The amount of transcripts of *LjeIF-4A* was used for normalization. As the result of quantitative real time RT-PCR, the expression level of *LjNSG1* in root nodules was almost 20 times higher than that in roots (Fig. 8).

Expression of sym- and nonsym-Hb genes in mycorrhiza

The expression of sym-Hb genes *LjLb1*, *2* and *3* could not be detected in the roots colonized by mycorrhizal fungi *Glomus* sp. R10 (Fig. 7). Any amplified DNA fragments could not be detected even though the number of amplification cycles in PCR was increased up to 35. This indicates that the infection of mycorrhizal fungi does not induce the expression of sym-Hb genes in *L. japonicus*.

The expression of nonsym-Hb gene *LjNSG1* was repressed in the roots colonized by mycorrhizal fungi *Glomus* sp. R-10 (Fig. 7). The expression level of *LjNSG1* in the roots detached nodules was estimated to be almost the same as that in the roots without any infection (data not shown). Thus, the repression of *LjNSG1* is expected to be a response to the infection of *Glomus* sp. The amount of transcripts of *LjNSG1* was compared between RNA from roots colonized by *Glomus* sp. and from roots without any infection by quantitative real time RT-PCR (Fig. 8). The amount of transcripts of *LjeIF-4A* was used for normalization. The expression level of *LjNSG1* in the roots colonized by *Glomus* sp. was 10 times less than that in the roots without colonization.

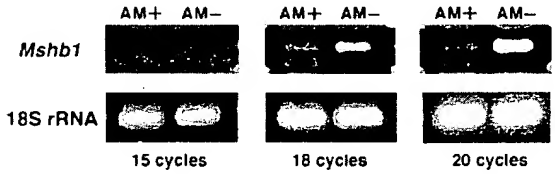

Fig. 9 Expression of nonsym-Hb gene in mycorrhiza of *Medicago sativa*. Total RNAs were isolated from roots with (AM+) or without (AM−) colonization of *Glomus* sp. R-10. RT-PCR was performed using specific primers for *Mhb1*, nonsym-Hb gene of *M. sativa*. Ribosomal RNA of *M. sativa* was used as a positive control.

To confirm whether repression on nonsym-Hb gene expression is specific to *L. japonicus* or not, the expression of *Mhb1*, nonsym-Hb gene of *Medicago sativa*, was analyzed by RT-PCR (Fig. 9). The expression of *Mhb1* was also repressed in the mycorrhiza of *Medicago sativa*. The expression level of *Mhb1* in mycorrhiza was estimated to be eight times less than that in the normal roots by quantitative RT-PCR (data not shown). These results suggest that mycorrhizal fungi control the expression of the nonsym-Hb gene during the establishment of symbiosis with leguminous plants.

Discussion

Lbs, sym-Hbs in legumes, are the most abundant soluble proteins in root nodules and their genes belong to multigene families (Brisson and Verma 1982, Appleby 1992). At least three sym-Hb genes, *LjLb1*, *LjLb2* and *LjLb3*, were identified in the genome of *L. japonicus*. The sym-Hb promoters contain

two motifs that have been shown to be critical for nodule specific expression (the nodulin motif, AAAGAT and CTCTT separated by six or seven nucleotides). The nodulin motifs of the three *LjLb* genes are slightly different each other (Fig. 2), whereas no significant differences were detected by RT-PCR in the quantity of mRNA in each gene in the root nodules (Fig. 5). *Pisum sativum* has at least five genes for sym-Hbs. The quantity of mRNA in each of the five genes was different in the nodules (Kawashima et al. 2001). Expression of the *PsLb5-10* gene was the highest, and that of *PsLb120-8* was the lowest. The nodule morphology of *L. japonicus* is the determinate type and that of *P. sativum* is the indeterminate type. The difference in the relative amounts of sym-Hb transcripts in the nodules will be due to the difference of nodule type between *P. sativum* and *L. japonicus*

Although sym-Hb has been thought to be a representative late nodulin, Heidstra et al. (1997) reported that the expression of sym-Hb gene of *Vicia sativa*, *VsLb1*, was induced within 1 h by a Nod-factor addition, unlike other late nodulin genes. In mature plants of *L. japonicus*, the expression of the three sym-Hb genes could be detected only in the root nodules the as same as sym-Hb genes in other legumes. Surprisingly, the transcripts of the three sym-Hb genes were detected in young shoots (3 d after germination) independent of Nod factor addition or *M. loti* infection, even though the expression was very low compared with that in nodules (Fig. 6). This result suggests that sym-Hb has another unknown function other than being the oxygen transporter for microsymbionts.

VfLb29, one of the sym-Hb genes of *Vicia faba*, is known to be induced both in root nodules and in roots colonized by the arbuscular mycorrhiza (AM) fungus *Glomus fasciculatum*, and other sym-Hb genes (*LfLbB, VfLbK, VfLb49*) were induced only in root nodules (Frühling et al. 1997). The possible function of VfLb29 was to supply oxygen for the respiration of the microsymbiont. The three sym-Hb genes in *L. japonicus* were not induced in the roots colonized by *Glomus* sp. R-10 (Fig. 7). The well-developed hyphae on the surface of the roots might transport sufficient oxygen to internal hyphae so that *Glomus* sp. R-10 does not need the help of sym-Hb in the host plants, and/or the control of expression might be quite different among *VfLb29* and sym-Hb genes in *L. japonicus*.

A gene that encodes nonsym-Hb was isolated from genomic library of *L. japonicus* and was referred to as *LjNSG1*. In *C. glauca* that establishes symbiotic nitrogen-fixing associations with the actinomycete *Frankia*, expression of the nonsym-Hb gene (*cashb-nonsym*) was not detectable in nodules (Jacobsen-Lyon et al. 1995). On the other hand, the nonsym-Hb gene (*Soyhb*) of *G. max* was expressed in root nodules and the expression level was slightly higher compared with that in shoots, leaves, flower buds and root (Andersson et al. 1996). *LjNSG1* was expressed in all of the tissues tested and the expression level was estimated to be 20-fold higher in nodules compared with that in roots by quantitative RT-PCR (Fig. 5, 8). The sequence of 5'-upstream region of *LjNSG1* contains the nodulin motif-like sequences (CACGAT-6 bases spacer-CTCTT), which resemble the nodulin motif sequences (CAAGAT-6 bases spacer-CTCTT) of sym-Hb genes of *C. glauca*. The promoter analysis of nonsym-Hb of *Trema tomentosa*, a kind of elm tree and non-nodulating plant, was reported by Andersson et al. (1997). The nodulin motif-like sequences (AAGGAG-4 bases spacer-CTCTC) and consensus sequence of nonsym-Hb (GAAGAG) were also found in the promoter of the nonsym-Hb gene of *T. tomentosa* (Fig. 2), and it was presumed that both AAGGAG and GAAGAG sequences were critical for the nonsym-Hb expression in nodules and roots of transgenic *Lotus corniculatus*. Both the consensus sequence of nonsym-Hb like sequence (GAAGGG) and nodulin motif like sequences (CACGAT-4 bases spacer-CTCTT) were found on the promoter region of *LjNSG1*. These sequences might be critical for the enhanced expression of *LjNSG1* in root nodules of *L. japonicus*.

There are several reports about expression and induction of nonsym-Hb gene in plants. The nonsym-Hb gene of soybean was expressed in many tissues but at different levels (Andersson et al. 1996). Accumulation of mRNA from the nonsym-Hb gene or the nonsym-Hb protein was detected under hypoxia in barley (Taylor et al. 1994) and in rice plants (Lira-Ruan et al. 2001). The nonsym-Hb gene in barley is shown to be involved in ATP metabolism under hypoxia (Sowa et al. 1998). The expression profile of the two nonsym-Hb genes of *A. thaliana*, *AHB1* and *AHB2*, is also well documented. The expression of *AHB1* was induced by hypoxia in roots and rosettes, whereas that of *AHB2* was induced by low temperature (Trevaskis et al. 1997). Wang et al. (2000) revealed that *AHB1* also responds to the level of nitrate using the microarray of *A. thaliana*. *Mhb1*, the nonsym-Hb gene of *M. sativa*, was also inducible by hypoxia but not by low temperature. The level of the transcript of *Mhb1* was increased at the G2/M boundary in a synchronized suspension culture (Seregélyes et al. 2000).

In this report, we focused on the response of Hb genes to microsymbionts and showed that the expression of the nonsym-Hb genes (*LjNSG1* and *Mhb1*) was repressed in mycorrhiza both in *L. japonicus* and *M. sativa* (Fig. 7, 8, 9). We do not have a reliable explanation for this observation, because nonsym-Hb is expected to have multiple roles in different plant tissues and to be involved in several metabolic pathways as described by Arredondo-Peter et al. (1998). Nonsym-Hb will be able to bind to small molecules known to be ligands of hemoglobins, e.g. O_2, CO and NO. NO is now known to be a signal for activation of the plant defense system (Durner et al. 1998, Klessig et al. 2000). Many investigators have reported that the AM-symbiosis reduces root disease caused by several soil-born pathogens (Davis and Menge 1980, Bärtschi et al. 1981, Mark and Cassells 1996, Murphy et al. 2000, Norman and Hooker 2000). Recently, Pozo et al. (2002) reported that *Glomus mosseae* was effective in reducing disease symptoms caused by *Phytophthora parasitica* infection on tomato roots

and that new isoforms of defense-related proteins (chitinase, chitosanase and β-1,3-glucanase) were induced in mycorrhiza. Although it is not clear whether nonsym-Hb is involved in the plant defense system directly, repression of the nonsym-Hb gene might facilitate the activation of plant defense system if nonsym-Hb plays a role as a NO scavenger in a normal plant cell. Further investigation on biochemical characterization will be required to understand the function of plant hemoglobins on symbiosis and the defense system.

Materials and Methods

Plants and microorganisms

Lotus japonicus MG20 (Kawaguchi 2000) and *Medicago sativa* (Yukijirushi Shubyo Co. Ltd.) were used as the host plants. *Mesorhizobium loti* MAFF303099 (Saeki and Kouchi 2000, Kaneko et al. 2000) and *Glomus* sp. R-10 (Dr. Kinkon, Idemitsu Kosan Co. Ltd., Japan) were used as the symbiotic partner.

Nodulation and mycorrhization

For nodulation, the seedlings of *L. japonicus* were inoculated with *M. loti* 3 d after germination The inoculated seedlings were transferred onto agar plates and were grown under a regime of 16 h light/8 h dark at 25°C. The nodules for RNA isolation were harvested at 6 weeks after inoculation. For mycorrhization, the inoculum and growth conditions were as described previously (Solaiman et al. 2000). The mycorrhizal colonization was confirmed by microscopy at 6 weeks after cultivation, and the whole root was used for RNA isolation.

Isolation of total DNA and RNA from plants

Genomic DNA of *L. japonicus* was isolated by extraction with CTAB (hexadecyltrimethyl ammonium bromide) and total RNA was isolated using a phenol-SDS procedure according to the method of Suzuki et al. (1997).

PCR for amplification and identification of genomic hemoglobin clones of L. japonicus

The primer pairs for amplification of sym-Hb genes were designed based on the sequences of leghemoglobin homologues (AF000390, AF000405, AF000406, AV418770) found in the EST libraries of root nodules of *L. japonicus* reported by Szczyglowski et al. (1997) and Asamizu et al. (2000): LjLbF1, GGCATATGAAACATTCAAGAA; LjLbR1, AACCATTGTGACTATATATTTCG, LjN77F1, GCAATAGTGTTCTGTTCTACACC; LjN77R1, GCGAAACCAGAAACATC. For screening of the 3-D pool of the genomic library of *L. japonicus*, other primer sets were used. The sequences of the primers were as follows: LjLbF2, AAAGACATGTTCTCCTTTCT; LjLb1R, TGGAATACATATGGTTATAG; LjLb2R, GCTATAAATTTCAACTTCAG; LjLb3R, AACAAGTTTTATTGGAGTAC. The primer pairs for the nonsym-Hb gene were designed based on the sequence of a nonsym-Hb homologue (AV413959) found in ESTs of seedlings of *L. japonicus* constructed by Asamizu et al. (2000): LjNSGF1, TTCTCACTTCACTTCCATCGC; LjNSGF2, TTGGTCAAGTCATGGAGCG; LjNSGR1, TCACAGTGACTTTTCCAGCG; LjNSGR2, AGACAGACATGGCATGAGGC. These primers for nonsym-Hb gene were also used for screening of the 3-D pool of the genomic library. PCR was performed in 30 cycles (94°C for 30 s, 55°C for 30 s, 72°C for 30 s). The PCR products were analyzed on 0.7% or 1.0% agarose gel depending on desired resolution.

Southern hybridization

The genomic DNA of *L. japonicus* was purified by CsCl-EtBr dye buoyant density gradient centrifugation and was digested with restriction endonucleases, *Pst*I, *Hin*dIII, *Bam*HI and *Eco*RI. The digests (approx. 4 μg) were separated on a 0.7% agarose gel, and then transferred to a nylon filter (Immobilon-Ny+; Millipore). Hybridization was carried out by the same conditions described in Suzuki et al. (2001). DNA probes for sym- and nonsym-Hb genes were obtained by PCR using corresponding TAC clone as a template. The sequences of primers used for amplification for each gene were as follows: AAAGACATGTTCTCCTTTCT and TGGAATACATATGGTTATAG for *LjLb1*; TAGAAAGATCTGTTACCGGT and GCCATTATGGCTATAAATTTCA for *LjLb2*; CAATGTTCCTATTGTCCACT and CTGCCAAACAAAGAATAGCA for *LjLb3*; TTCTCACTTCACTTCCATCGC and TCACAGTGACTTTTCCAGCC for *LjNSG1*. By using these primer sets, DNA fragments ranging from the second exon to 3′-UTR of *LjLb1*, ranging from the third intron to 3′-UTR of *LjLb2*, ranging from the first exon to the third exon of *LjNSG1*, and a part of the third intron of *LjLb3* were obtained, respectively. The PCR products were labeled with ^{32}P using Multiprime DNA labeling system (Amersham) and used as probes.

DNA sequencing and sequence analysis

Sequencing reactions were carried out according to the manufacture's instruction using the BigDye Terminator Cycle Sequencing FS Reaction Kit (Applied Biosystems) and the sequences were read automatically using ABI PRISM 310 and 377, Applied Biosystems). Analyses of sequences and multiple alignment of putative amino acid sequences were performed by the GeneWorks (Intelligenetics) and MacVector programs (Genetics Computer Group).

Analysis of expression of Hb genes by RT-PCR

The expression of each Hb gene was detected and estimated by RT-PCR (OneStep RT-PCR; Qiagen) using specific primer pairs as follows: LjLbF2 and LjLb1R for *LjLb1*; LjLbF2 and LjLb2R for *LjLb2*; LjLbF2 and LjLb3R for *LjLb3*; LjNSGF1 and LjNSGR2 for *LjNSG1*. For Fig. 6, first strand cDNA was synthesized by SuperScriptII (Invitrogen) and *ExTaq* (TaKaRa) was used for amplification. The degenerate primer pair for the three *LjLbs* was also used in Figs 6 and 7: PEALLLb-F, TCTGGRCCYAMGCAYAGTC; PEALLLb-R CRTCRCGWGTCAGTSCAAAA. For detection of the gene expression of *Mhb1* (nonsym-Hb gene of *Medicago sativa*, AF172172), the following primers were used: Mhb1F, CATGACATGTGAATCAGCCG; Mhb1R, CCAACTGATCATAAGCTTCTC. The genes for initiation factor 4A3 of *L. japonicus*, *LjeIF-4A* and 18S rRNA of *M. sativa* were used for positive control. The primers for detection of 18S rRNA of *M. sativa* were CCTAGTAAGCGCGAGTCATC and CATTCAATCGGTAGGAGCGA. The amplified products were separated on 1% agarose gel electrophoresis.

To quantify the relative amount of transcripts derived from each Hb gene, real time RT-PCR was employed. DNase I-treated total RNA (100 ng) was used as a template in a 50 μl of reaction mixture containing 25 μl of SYBR Green PCR master mix (Applied Biosystems), 1 μl of RNase inhibitor (40 U μl^{-1}), 0.25 μl of Superscript II-RT (Invitrogen, 200 U μl^{-1}), and 2 μl of each primer (5 μM). RT reaction and PCR were performed using 7700 Sequence Detection System (Applied Biosystems). PCR was carried out 40 cycles of 95°C for 15 s and 60°C for 1 min. The sequences of primers for real time RT-PCR were as follows: CCTTTGGAGGAGAACCCCAA and GAGCTGCTGATTCACAAGTCATG for *LjNSG1*; AGAGGGTTTAAAGATCAAAT and ATGTCAATTCATCACGTTTT for *LjeIF-4A*. These sequences were designed with Primer Express program (Applied Biosystems). The primer pairs suitable for independent detection of the three *LjLb*s could

not be designed by this program. The gene for initiation factor 4A3 of *L. japonicus*, *LjeIF-4A*, was used as an internal standard for RT-PCR and was used for normalization of the results of quantitative RT-PCR.

Acknowledgments

The authors thank Dr. Motoshi Suzuki and Mr. Akihiko Narutaki of Idemitsu Kosan Co, Ltd., Tokyo, Japan for supplying spores of *Glomus* sp. R-10. This work was supported in part by Special Coordination Funds for Promoting Science and Technology, and by a Grant-in-Aid for Scientific Research (C) (no. 13640653) from Japan Society for the Promotion of Science (JSPS).

References

Andersson, C.R., Jensen, E.O., Llewellyn, D.J., Dennis, E. and Peacock, W.J. (1996) A new hemoglobin gene from soybean: a role for hemoglobin in all plants. *Proc. Natl. Acad. Sci. USA*. 93: 5682–5687

Andersson, C.R., Llewellyn, D.J., Peacock, W.J. and Dennis, E.S. (1997) Cell-specific expression of the promoters of two nonlegume hemoglobin genes in a transgenic legume, *Lotus corniculatus*. *Plant Physiol.* 113: 45–57.

Appleby, C.A. (1992) The origin and functions of haemoglobin in plants. *Sci. Prog.* 76: 365–398.

Arredondo-Peter, R., Hargrove, M.S., Moran, J.F., Sarath, G. and Klucas, R. (1998) Plant hemoglobins. *Plant Physiol.* 118: 1121–1125.

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S. (2000) Generation of 7137 non-redundant expressed sequence tags from a legume, *Lotus japonicus*. *DNA Res.* 7: 127–130.

Bärtschi, H., Gianinazzi-Pearson, V. and Vegh, I. (1981) Vesicular-arbuscular mycorrhizal formation and root rot disease (*Phytophthora cinnamomi*) development in *Chamaecyparis lawsoniana*. *Phytopathology* 102: 213–218.

Brisson, N. and Verma, D.P.S. (1982) Soybean leghemoglobin gene groups: normal, pseudo, and truncated genes. *Proc. Natl. Acad. Sci. USA* 79: 4055–4059.

Davis, R.M. and Menge, J.A. (1980) Influence of *Glomus fasciculatus* and soil phosphorus on *Phytophthora* root rot of citrus. *Phytopathology* 70: 447–452.

Durner, J., Wendehenne, D. and Klessig, D.F. (1998) Defense gene induction in tobacco by nitric oxide, cyclic GMP, and cyclic ADP-ribose. *Proc. Natl. Acad. Sci. USA* 95: 10328–10333.

Frühling, M., Roussel, H., Gianinazzi-Pearson, V., Pühler, A. and Perlick, A. (1997) The *Vicia faba* leghemoglobin gene *VfLb29* is induced in root nodules and in roots colonized by the arbuscular mycorrhizal fungus Glonus fasciculatum. *Mol. Plant Microbe Interact.* 10: 124–131.

Heidstra, R., Nilsen, G., Martinez-Abarca, F., van Kammen, A. and Bisseling, T. (1997) Nod factor-induced expression of leghemoglobin to study the mechanism of NH_4NO_3 inhibition on root hair deformation. *Mol. Plant Microbe Interact.* 10: 215–220.

Jacobsen-Lyon, K., Jensen, E.Ø, Jørgensen, J.-E., Marcker, K.A., Peacock, W.J. and Dennis, E.S. (1995) Symbiotic and nonsymbiotic hemoglobin genes of *Casuarina glauca*. *Plant Cell* 7: 213–223.

Kaneko, K., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Hohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.* 7: 331–338.

Kapranov, P., de Bruijn, F.J. and Szczyglowski, K. (1997) Novel, highly expressed late nodulin gene (*LjNOD16*) from *Lotus japonicus*. *Plant Physiol.* 113: 1081–1090.

Kawaguchi, M. (2000) *Lotus japonicus* 'Miyakojima' MG-20: an early-flowering accession suitable for indoor handling. *J. Plant Res.* 113: 507–509.

Kawashima, K., Suganuma, N., Tamaoki, M. and Kouchi, H. (2001) Two types of pea leghemoglobin genes showing different O_2-binding affinities and distinct patterns of spatial expression in nodules. *Plant Physiol.* 125: 641–651.

Klessig, D., Durner, J., Noad, R., Navarre, D.A., Wendehenne, D., Kumar, D., Zhou, J.M., Shah, J., Zhang, S., Kachroo, P., Trifa, Y., Pontier, D., Lam, E. and Silva, H. (2000) Nitric oxide and salicylic acid signaling in plant defense. *Proc. Natl. Acad. Sci. USA* 97: 8849–8855.

Lira-Ruan, V., Sarath, G., Klucas, R.V. and Arredondo-Peter, R. (2001) Synthesis of hemoglobin in rice (*Oryza sativa* var. Jackson) plants growing in normal and stress conditions. *Plant Sci.* 161: 279–287.

Liu, Y.G., Shirano, Y., Fukaki, H., Yanai, Y., Tasaka, M., Tabata, S. and Shibata, D. (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. *Proc. Natl. Acad. Sci. USA* 96: 6535–6540.

Mark, G.L. and Cassells, A C. (1996) Genotype-dependence in the interaction between *Glomus fistulosum*, *Phytophthora fragariae* and the wild strawberry (*Fragaria vesca*). *Plant Soil* 185: 233–239.

Murphy, J.G., Rafferty, S.M. and Cassells, A.C. (2000) Stimulation of wild strawberry (*Fragaria vesca*) arbuscular mycorrhizas by addition of shellfish waste to the growth substrate: interaction between mycorrhization, substrate amendment and susceptibility to red core (*Phytophthora fragariae*). *Appl. Soil Ecol.* 15: 153–158.

Norman, J.R. and Hooker, J.E. (2000) Sporulation of *Phytophthora fragariae* shows greater stimulation by exudates of non-mycorrhizal than by mycorrhizal strawberry roots. *Mycol. Res.* 104: 1069–1073.

Pozo, M.J., Cordier, C., Dumas-Gaudot, E., Gianinazzi, S., Barea, J.M. and Azcón-Aguilar, C. (2002) Localized versus systemic effect of arbuscular mycorrhizal fungi on defence responses to *Phytophthora* infection in tomato plants. *J. Exp. Bot.* 53: 525–534.

Saeki, K. and Kouchi, H. (2000) The *Lotus* symbiont, *Mesorhizobium loti*: molecular genetic techniques and application. *J. Plant Res.* 113: 457–465.

Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. and Tabata, S. (2001) Structural analysis of a *Lotus japonicus* genome. I. Sequence features and mapping of fifty-six TAC clones which cover the 5.4 Mbyte regions of the genome. *DNA Res.* 8: 311–318

Seregélyes, C., Mustárdy, L., Ayaydin, F., Sass, L., Kovács, L., Endre, G., Lukács, N., Kovács, I., Vass, I., Kiss, G.B., Horváth, G.V. and Dudits, D. (2000) Nuclear localization of a hypoxia-inducible novel non-symbiotic hemoglobin in cultured alfalfa cells. *FEBS Lett.* 482: 125–130.

Solaiman, M.Z., Snoo, K., Kawaguchi, M., Imaizumi-Anraku, H., Akao, S., Tanaka, A. and Obata, H. (2000) Characterization of mycorrhizas formed by *Glomus* sp. on roots of hypernodulating mutants of *Lotus japonicus*. *J. Plant Res.* 113: 443–448.

Sowa, A.W., Duff, S.M.G., Guy, P.A. and Hill, R.D. (1998) Altering hemoglobin levels changes energy status in maize calls under hypoxia. *Proc. Natl. Acad. Sci. USA* 95: 10317–10321.

Suzuki, A., Suzuki, T., Tanabe, F., Toki, S., Washida, H., Wu, C.Y. and Takaiwa, F. (1997) Cloning and expression of five myb-related genes from rice seed. *Gene* 198: 393–398.

Suzuki, A., Kobayashi, F., Abe, M., Uchiumi, T. and Higashi, S. (2001) Cloning and expression of a down-regulated gene (*TrEnodDR1*) of white clover responded by the nod genes derived from *Rhizobium leguminosarum* bv. *trifolii* strain 4S. *Gene* 266: 77–84.

Szczyglowski, K., Hamburger, D., Kapranov, P. and de Bruijn, F.J. (1997) Construction of a *Lotus japonicus* late nodulin expressed sequence tag library and identification of novel nodule-specific genes. *Plant Physiol.* 114: 1335–1346.

Trevaskis, B., Watts, R.A., Andersson, C.R., Llewellyn, D.J., Hargrove, M.S., Olson, J.S., Dennis, E.S. and Peacock, W.J. (1997) Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins. *Proc. Natl. Acad. Sci. USA* 94: 12230–12234.

Taylor, E.R., Nie, X.Z., MacGregor, A.W. and Hill, R.D. (1994) A cereal haemoglobin gene is expressed in seed and root tissues under anaerobic conditions. *Plant Mol. Biol.* 24: 853–862.

Wang, R., Guegler, K., LaBrie, S.T. and Crawford, N.M. (2000) Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell* 12: 1491–1509.

(Received April 30, 2002; Accepted September 9, 2002)

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimica et Biophysica Acta 1621 (2003) 299–305
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Short sequence-paper

Molecular cloning and characterization of cDNAs encoding hemoglobin from wheat (*Triticum aestivum*) and potato (*Solanum tuberosum*)☆

Knud Larsen*

Department of Crop Physiology and Soil Science, Danish Institute of Agricultural Sciences, P.O. Box 50, DK-8830 Tjele, Denmark

Received 14 January 2003; received in revised form 15 April 2003; accepted 25 April 2003

Abstract

Hemoglobins (Hbs) are heme proteins encountered in all five kingdoms of living organisms. In plants, two different classes of Hbs have been identified: nonsymbiotic (class I) from both monocot and dicot species and symbiotic (class II) Hbs from nitrogen-fixing plants. This work reports the cloning and analysis of three nonsymbiotic Hb genes from wheat (*Triticum aestivum*) and potato (*Solanum tuberosum*). The Hb cDNAs were amplified by reverse transcriptase polymerase chain reaction (RT-PCR) using consensus oligonucleotide primers for nonsymbiotic Hbs.

A wheat Hb cDNA (*TaHb1*) was isolated and shows a very high similarity to nonsymbiotic Hbs from *Hordeum vulgare* (98%) and *Zea mays* (83%). Another wheat Hb cDNA, designated *TaHb2*, exhibited strong similarity to truncated bacterial Hbs, the so-called 2-on-2 Hbs. In addition, a third Hb was cloned from potato, *StHb*. Expression analysis by RT-PCR demonstrated a very high expression level of the *TaHb1* gene only in wheat roots. In contrast, the other wheat hemoglobin gene, *TaHb2*, was demonstrated to be constitutively expressed although differences in expression level in different tissues were observed. The expression of the *TaHb1* gene is induced in wheat roots exposed to microaerobic conditions. The potato Hb gene, *StHb*, was highly expressed in roots and also in tubers and stem tissue although at much reduced levels.

Keywords: cDNA; Hemoglobin; Nonsymbiotic; Potato; *Solanum tuberosum*; *Triticum aestivum*; Wheat

Hemoglobins (Hbs) are heme proteins encountered in all five kingdoms of living organisms whose primary function is the binding and transporting oxygen and other gaseous ligands [1,2]. Hbs function by means of reversible binding in the storage and transportation of oxygen and by facilitating its diffusion. In plants, three different types of Hbs have been described: symbiotic, nonsymbiotic and truncated Hbs. The symbiotic Hbs are found in leguminous plants and in actinorhizal plants existing in symbiosis with N_2-fixing organisms [3]. However, symbiotic-like Hbs are also found in plants as *Cichorium* and display a close sequence similarity to symbiotic Hbs [4]. The symbiotic Hbs function as oxygen buffers and carriers and provide O_2 to the respirating micro-symbionts actively fixing N_2 in tissues. Symbiotic Hbs also prevent inhibition of the nitrogenase complex by binding O_2 tightly and facilitating its diffusion to N_2-fixing bacteroids.

Nonsymbiotic Hbs are ubiquitous in plant species and have been divided into two classes: Class II with oxygen-binding properties like symbiotic Hbs and class I with significantly different oxygen-binding properties. The first nonsymbiotic Hb was identified in barley and the identity established from its similarity to other nonlegume Hbs [5]. Since then nonsymbiotic Hbs have been found in virtually all plant species examined including liverworts (*Marchantia polymorpha*) [6], mosses (*Physcomitrella patens*) [7], the monocots rice, maize and teosinte [8–10], and the dicots soybean, *Arabidopsis* [11,12]. The nonsymbiotic Hbs dis-

Abbreviations: GAPDH, glyceraldehyde-3-phosphate dehydrogenase; Hb, hemoglobin; ORF, open reading frame; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase PCR; UTR, untranslated region

☆ The cDNA sequences for *TaHb1*, *TaHb2* and *StHb* have been submitted to DDBJ/EMBL/GenBank under the accesssion numbers AY151390, AY151391 and AY151389, respectively.

* Present address: Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, P.O. Box 50, DK-8830 Tjele, Denmark. Tel.: +45-89991827; fax: +45-89991839, +45-89991300.

E-mail address: Knud.Larsen@agrsci.dk (K. Larsen).

0304-4165/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0304-4165(03)00094-1

play expression patterns different from the symbiotic Hbs and are generally expressed at low levels and found in a variety of tissues including roots, stems and leaves of both monocots and dicots [13]. The expression patterns of non-symbiotic *Hb* genes varies significantly among different species, and often the highest levels of expression are detected in metabolically active or stressed organs/tissues. A number of nonsymbiotic Hbs belonging to class I have been shown to be hypoxia-inducible in plants such as barley [14], maize [15], rice [16] and alfalfa [17,18], and are often referred to as stress-induced Hb.

The function of nonsymbiotic Hb in plants has recently been reviewed by Ross [19] and Dordas et al. [20]. Non-symbiotic Hbs might be oxygen-transporter proteins that scavenge oxygen during hypoxia and supply it for aerobic respiration, or Hbs can act as terminal NADH oxidases that facilitate glycolytic formation of ATP under hypoxic conditions. A third possible function for Hbs could be as oxygen-sensing proteins that activate other regulatory proteins. It seems most plausible that nonsymbiotic Hbs are involved in maintaining the energy status of plant cells exposed to low oxygen concentrations. Dordas et al. [20] propose a function of hypoxia stress-induced Hb in modulation of nitric oxide in the plant cell. Recently the first crystal structure of a nonsymbiotic rice Hb (Hb1) was determined [21]. However, the functional relevance of the quaternary structure is still unclear.

Recently a third group of Hbs, originally identified in prokaryotes, protozoa and algae, has also been demonstrated to be present in plants. A so-called truncated Hb was isolated from *Arabidopsis thaliana* [22]. The *Arabidopsis* GLB3 protein has a three-dimensional structure similar to the characteristic 2-on-2 arrangement α-helices, distinct from the 3-on-3 arrangement of the standard globin fold [23]. Expressed sequence tags with high similarity to *A. thaliana GLB3* have been identified in both dicots and monocots including barley and maize. *GLB3* is expressed throughout the plant and does not respond to hypoxia or cold stress and, hence, *GLB3* expression is regulated differently from the two other *Arabidopsis* Hbs.

In the present study we have undertaken the cloning and analysis of wheat and potato Hb genes. The spatial expression pattern of the two wheat Hb genes was investigated in response to limiting oxygen stress.

Tissues of wheat (*Triticum aestivum* var. Complet) used in the reported experiments were collected from field-grown plants (Danish Institute of Agricultural Sciences, Tjele, Denmark) or grown in a hydroponic culture. Potato (*Solanum tuberosum* cv 68-DCU-28, a tetraploid) plants grown under green-house conditions were also used as plant source for this experiment. Plant material used for reverse transcriptase polymerase chain reaction (RT-PCR) was immediately frozen in liquid nitrogen after harvest and stored at − 80 °C. Nucleic acids were extracted and purified from deep-frozen plant material. Total RNA was extracted by employing the RNeasy Plant Mini Kit from Qiagen according to the manufacturer's instructions. RNA concentrations were determined by UV spectrometry.

Total RNA was isolated from wheat seedlings and potato stems as described above. After treatment with RNase-free DNase, 5 μg of total RNA was used in an RT-PCR one-tube reaction employing a RobusT RT-PCR Kit (Finnzymes). Oligonucleotide primers were derived from putative Hb sequences in the EST databases found in TIGR (http://www.tigr.org/tdb/). The RT-PCR reaction mix contained: 5 μg total RNA, 1.5 mM $MgCl_2$, 0.2 mM dNTP, 0.5 μM of each primer (TaHb1-F: 5′-GGGAGGGAGGAAGCCATGT-CTGCCGCGGAG-3′ and TaHb1-R: 5′-ACAATGGATGA-GCTATTCAGAGGGCTTCAT-3′ for cloning of the wheat *Hb1* gene, TaHb2-F: 5′-GGCGCTGCAGGAACGATG-CAGTCGCTGCAGG-3′ and TaHb2-R: 5′-CCGG-TTAAGTTTGAGCGGTTACTCGGCTGG-3′ for cloning of the wheat *Hb2* gene and StHb-F: 5′-AAAGTTTGATCA-CAATCATCATGAGTAGCTTT-3′ and StHb-R: 5′-GAGCCGATAGTCCATCGGAATCAGTCC-3′ for cloning of the potato *Hb* gene), 20 units of DyNAzyme EXT DNA polymerase (Finnzymes), 5 units of AMV reverse transcriptase (Finnzymes) and 20 units of RNase inhibitor (Roche) in a total volume of 50 μl. Synthesis of cDNA, inactivation of AMV reverse transcriptase and denaturation of cDNA/mRNA hybrids were performed by one cycle of: 50 °C for 60 min followed by 94 °C for 2 min. PCR amplification was carried out in the total volume by the use of the following program: 30 cycles of 45 °C for 30 s followed by 72 °C for 1 min 30 s. The PCR program was concluded by an extension at 72 °C for 7 min. Ten-microliter PCR amplification products were applied on a 1% agarose gel and visualized after electrophoresis by ethidium bromide staining. Fluorescent bands of sizes of approximately 500 bp for *TaHb1*, 550 bp for *TaHb2* and 650 bp for *StHb* were cut out and eluted using the Qiaquick Gel Extraction kit from Qiagen. The eluted PCR products were cloned into the pCR TOPO 2.1 vector (Invitrogen) and sequenced in both directions employing the dideoxy chain termination method using BigDye terminator cycle sequencing kit with AmpliTaq DNA polymerase FS (PE Applied Biosystems). The sequencing analysis was performed on an automated DNA sequencer (ABI PRISM™ Genetic Analyzer Model 373/377) (PE Applied Biosystems) using facilities within the Institute of Molecular and Structural Biology, University of Aarhus, Denmark. All DNA fragments were sequenced in both directions.

The identities of TaHb1, TaHb2 and StHb were demonstrated by comparison of the deduced polypeptide sequences to other isolated plant Hbs. The wheat hemoglobin1 (Hb1) cDNA, *TaHb1*, consists of a 516-bp fragment with the translational start site of the major open reading frame (ORF) at nucleotide 16 and the TAG stop site at nucleotide 504. The deduced amino acid sequence of TaHb1 is shown in Fig. 1. Comparison of the length of the wheat Hb1 with other Hb genes indicates that *TaHb1* is not full-length and that both 5′UTR and 3′UTR are incomplete. The transla-

```
PotatoHb.(d)     ---MSS----------FSEEQEALVVKSWGSMKKDAGEWGLKFFLKIFEIAPSAKKMFSF 47
TomatoHb.(d)     ---MSS----------FSEEQEALVVKSWGSMKKDAGEWGLKFFLKIFEIAPSAKKMFSF 47
CasuarHb2.(d)    ---MSTLEGR----G-FTEEQEALVVKSWSAMKPNAGELGLKFFLKIFEIAPSAQKLFSF 52
SoybeanHb.(d)    ---MTTTLER----G-FSEEQEALVVKSWNVMKKNSGELGLKFFLKIFEIAPSAQKLFSF 52
CottonHb.(d)     ---MATYEGK----V-FTEEQEALVVKSWTVMKKKTAELGLKFFLKIFEIAPSAKKLFSF 52
TremaHb.(d)      ---MSSSEVD----KVFTEEQEALVVKSWAVMKKNSAELGLKFFLKIFEIAPSANNLFSY 53
ParaspHb.(d)     ---MSSSEVN----KVFTEEQEALVVKAWAVMKKNSAELGLQFFLKIFEIAPSAKNLFSY 53
AraHb1.(d)       ---MES-EGK----IVFTEEQEALVVKSWSVMKKNSAELGLKLFIKIFEIAPTTKKMFSF 52
WheatHb1.(m)     ---MSAAERA----VVFSEEKDALVLKSWAIMKKDSANLGLRFFLKIFEIAPSARQMFPF 53
BarleyHb.(m)     ---MSAAEGA----VVFSEEKEALVLKSWAIMKKDSANLGLRFFLKIFEIAPSARQMFPF 53
MaizeHb.(m)      MALAEADDGA----VVFGEEQEALVLKSWAVMKKDAANLGLRFFLKVFEIAPSAEQMFSF 56
RiceHb1.(m)      MALVEDNNAV---AVSFSEEQEALVLKSWAILKKDSANIALRFFLKIFEVAPSASQMFSF 57
RiceHb2.(m)      MALVEGNNGVSGGAVSFSEEQEALVLKSWAIMKKDSANIGLRFFLKIFEVAPSASQMFSF 60
                    :             * **::***:*:*  :*  ::: :*::*:*:**:**:: ::*::

PotatoHb.(d)     LKDSNVPLDQNPKLKVHAKSILVMTCEAAVQLRKAGKVVVRDSTLKKIGATHFKYGVADE 107
TomatoHb.(d)     LKDSNVPLDQNPKLKIHAKSVLVMTCEAAVQLRKAGKVVVRDSTLKKIGATHFKYGVVDE 107
CasuarHb2.(d)    LKDSNVPLERNPKLKSHAMSVFLMTCESAVQLRKAGKVTVRESSLKKLGASHFKHGVADE 112
SoybeanHb.(d)    LRDSTVPLEQNPKLKPHAVSVFVMTCDSAVQLRKAGKVTVRESNLKKLGATHFRTGVANE 112
CottonHb.(d)     LRDSNVPLEQNTKLKPHAMSVFVMTCESAVQLRKAGKVTVRESNLKKLGATHFKYGVVDE 112
TremaHb.(d)      LKDSPIPLEQNPKLKPHAMTVFVMTCESAVQFRKAGKVTVRESNLKRIGAIHFKNGVVHE 113
ParaspHb.(d)     LKDSPVPLEQNPKLKPHATTVFVMTCESAVQLRKAGKVTVKESDLKRIGAIHFKTGVVNE 113
AraHb1.(d)       LRDSPIPAEQNPKLKPHAMSVFVMCCESAVQLRKTGKVTVRETTLKRLGASHSKYGVVDE 112
WheatHb1.(m)     LRDSDVPLETNPKLKTHAVSVFVMTCEAAAQLRKAGKITVRETTLKRLGGTHLKYGVADG 113
BarleyHb.(m)     LRDSDVPLETNPKLKTHAVSVFVMTCEAAAQLRKAGKITVRETTLKRLGGTHLKYGVADG 113
MaizeHb.(m)      LRDSDVPLEKNPKLKTHAMSVFVMTCEAAAQLRKAGKVTVRETTLKRLGATHLRYGVADG 116
RiceHb1.(m)      LRNSDVPLEKNPKLKTHAMSVFVMTCEAAAQLRKAGKVTVRDTTLKRLGATHLKYGVGDA 117
RiceHb2.(m)      LRNSDVPLEKNPKLKTHAMSVFVMTCEAAAQLRKAGKVTVRDTTLKRLGATHFKYGVGDA 120
                 *::* :*:: *:*** ** :::: *: *:*:**:**::*:: :**::*::*:: **

PotatoHb.(d)     HFEVTKYALLETIKEASP-EMWSVEMKNAWGEAYDQLVSAIKTEMK------ 152
TomatoHb.(d)     HFEVTKYALLETIKEASQ-EMWSVEMKNAWGEAYDQLVSAIKTEMK------ 152
CasuarHb2.(d)    HFEVTKFALLETIKEAVP-ETWSPEMKNAWGEAYDKLVAAIKLEMKPSS--- 160
SoybeanHb.(d)    HFEVTKFALLETIKEAVP-EMWSPAMKNAWGEAYDQLVDAIKSEMKPPSS-- 161
CottonHb.(d)     HFEVTKFALLETIKEAVP-DMWSDEMKNAWGEAYDRLVAAIKIEMKACSQAA 163
TremaHb.(d)      HFEVTRFALLETIKEAVP-EMWSPEMKNAWGEAYDQLVAAIKFEVKPSST-- 162
ParaspHb.(d)     HFEVTRFALLETIKEAVP-EMWSPEMKNAWGVAYDQLVAAIKFEMKPSST-- 162
AraHb1.(d)       HFEVAKYALLETIKEAVP-EMWSPEMKVAWGQAYDHLVAAIKAEMNLSN--- 160
WheatHb1.(m)     HFEVTRFALLETIKEALPADMWGPEMRNAWGEAYDQLVAAIKQEMKPSE--- 162
BarleyHb.(m)     HFEVTRFALLETIKEALPADMWGPEMRNAWGEAYDQLVAAIKQEMKPAE--- 162
MaizeHb.(m)      HFEVTGFALLETIKEALPADMWSLEMKKAWAEAYSQLVAAIKREMKPDA--- 165
RiceHb1.(m)      HFEVVKFALLDTIKEEVPADMWSPAMKSAWSEAYDHLVAAIKQEMKPAE--- 166
RiceHb2.(m)      HFEVTRFALLETIKEAVPVDMWSPAMKSAWSEAYNQLVAAIKQEMKPAE--- 169
                 ****: :***:****::: :*:::*::**::**::**:*** *::
```

Fig. 1. Multiple alignment of TaHb1 and StHb with hemoglobins from other plant species obtained by Clustal W. The numbers represent the position of the amino acids in the respective protein sequences. Identical amino acid residues in all sequences are indicated by asterisks. Semicolon indicates residues that are mostly conserved (identified in at least 8 out of 13 amino acids). Dashes within a sequence indicate gaps inserted to optimize alignment. The accession numbers of the sequences used in the comparison are: potatoHb, *S. tuberosum*, AY151389; tomatoHb, *L. esculentum*, AY026343; CasuarinaHb2, *C. glauca*, X53950; SoybeanHb, G. max, U47143; CottonHb, *G. hirsutum*, AF329368; TremaHb, *T. orientalis*, Z99635; ParasponiaHb, *P. andersonii*, U27194, U94998; WheatHb, *T. aestivum*, AY151390; BarleyHb, *H. vulgare*, U94968; MaizeHb, *Z. mays*, AF236080; RiceHb1, *O. sativa*, AF335504; RiceHb2, *O.sativa*, U76031. Alignments of sequences were performed using the Clustal W program on EBI WWW molecular biology server.

tional start site is in accordance with the plant consensus having a C at position +5. No putative poly(A^+) signal is observed due to lack of complete 3′UTR. The G+C content of the coding region is typical for a graminaceous monocot 52% [24], and it codes for a polypeptide of 162 amino acids with an estimated molecular mass of 18.1 kDa and p*I* of 8.7 (calculated using the sequence analysis tools of http://www.expasy.org/tools/dna.html).

The cDNA of the other isolated wheat Hb gene, *TaHb2*, is 541-bp long with the translational start site at nucleotide 16 and the TAA stop site at nucleotide 531. The ORF of *TaHb2* has a G+C content of 56% and it encodes a polypeptide of 171 amino acids with an estimated molecular mass of 19.5 kDa and p*I* of 5.6. The deduced amino acid sequence is shown in Fig. 2. The nomenclature 2-on-2 is referring to the arrangement of the α-helices in the three-dimensional structure of these Hbs.

The cDNA of the potato Hb gene (*StHb*) is 646-bp long with the translational start site at nucleotide 22 and the TAG stop at nucleotide 480. The ORF of *StHb* has a G+C content of 40% and it encodes a polypeptide of 152 amino acids with an estimated molecular mass of 17.1 kDa and p*I* of 8.9. The translational start of the *StHb* cDNA matches the plant consensus having an A at position −3. The lengths of most published Hb cDNAs are in the range of 600–900 bp and hence the TaHb1, TaHb2 and *StHb* cDNAs do not represent full-length cDNA clones. Obviously, both 5′UTR and 3′UTR sequences are missing.

Amino acid sequence similarity between TaHb1, StHb and other published plant Hbs was determined by the Clustal

```
TaHb2.2-on-2    MQSLQDKASEWSGVAAADAFAIDEVNVFEALGGTPQPFVDLSTNFYTRVYEDEEEWFPEI 60
HvHb.2-on2      MQSLQDKASEWSGVAAADAFAIDEVNVFEALGGTPQPFVDLSTNFYTRVYEDEEEWFREI 60
AtHb.2-on-2     MQSLQDKASVLSGVDQAEAFAIDESNLFDKLG--LQTFINLSTNFYTRVYDDEEEWFQSI 58
                ********* *** * ****** * *  **   * *  ********** ******  *

TaHb2.2-on-2    FSGSRKEDAIQNQYEFLVQRMGGPPLFSQRRGHPALIGRHRPFPVTHQAAERWLHHMQQA 120
HvHb.2-on-2     FSGSKKEDAIQNQYEFLVQRMGGPPLFSQRRGHPALIGRHRPFPVTHRAAERWLHHMQQA 120
AtHb.2-on-2     FSNSNKEDAIQNQYEFFVQRMGGPPLYSQRKGHPALIGRHRPFPVTHQAAERWLEHMQNA 118
                ** * ********** *********:***:**************** ****** *** *

TaHb2.2-on-2    LETTESINPDTKTKMMIFFRHTAYFLVAGNEMTRQTQSVP-----PCKHATSKPAE-    171
HvHb.2-on-2     LETTQSINPDTKTKMMNFFRHTAYFLVAGNEMTRQTQSVP-----PCKHATNKPAE-    171
AtHb.2-on-2     LDDSVDIDQDSKIKMMKFFRHTAFFLVAGNELKNQNEKPKHKPQCACKPAANKPAEE    175
                *     *  * * *** ****** *******   *          ** *  ****
```

Fig. 2. Alignment of 2-on-2 plant hemoglobins obtained by Clustal W. The numbers indicate the position of the amino acids in the respective protein sequences. Identical amino acid residues in all three sequences are indicated by an asterisk. Dashes within a sequence indicate gaps inserted to optimize alignment. The accession numbers of the sequences used in the comparison are: TaHb2.2-on-2, *T. aestivum*, AY151391; HvHb.2-on-2, *H. vulgare*, AF376063; AtHb.2-on-2, *A. thaliana*, AF376062. Alignment of sequences was performed using the Clustal W program on EBI WWW molecular biology server.

method. A substantial degree of identity between TaHb1, StHb and most other Hb was found. The encoded TaHb1 polypeptide exhibits significant sequence similarity to Hbs from different plants with highest identity to Hbs from *Hordeum vulgare* (98%) and *Zea mays* (83%). Generally, the highest similarity values are exhibited between TaHb1 and Hbs from other monocot species. The lowest amino acid identity between TaHb1 and other Hbs was, not surprisingly, to Hbs from dicot species as *Citrus unshiu* (64%) and *A. thaliana* (66%). The StHb amino acid sequence showed highest identity to Hbs from the dicots to tomato (*Lycopersicon esculaentum*), which like potato is a member of Solanaceae, and to *Gyssopium hirsutum* with values of 97% and 80%, respectively. TaHb2 exhibits high sequence similarity to 2-on-2 Hbs identified from different plants with highest identity (96%) to GLB3 from *H. vulgare* and GLB3 from *Z. mays* (88%) and also to *A. thaliana* 2-on-2 Hb (71%).

Multiple alignment of TaHb1, StHb and other plant nonsymbiotic Hb sequences (Fig. 1) revealed several regions of high homology. The highly conserved phenylalanine residue found in many known plant nonsymbiotic Hbs is also present in TaHb1 and StHb in positions F51 and F45, respectively. A cysteine residue found in nonsymbiotic Hbs is also present in TaHb1 (C79) and StHb (C73). With the exception of N at position 64, the TaHb1 contains the signature of plant Hbs (S-P-X-L-X-X-H-A-X-X-X-F). Sequence comparison also shows that wheat Hb1 contains distal (H70) and proximal (H105) histidine residues, as well as the P45, F51, F75 and F115 that are conserved in plant Hbs (Fig. 1). TaHb1 also contains a single Cys residue, C79, that is highly conserved in nonlegume Hbs. With the exception of N and L at positions 58 and 69, respectively, the StHb contains the consensus of plant Hbs (S-P-X-L-X-X-H-A-X-X-X-F) as well as the distal (H64) and proximal (H99) histidine residues, P39 and F109, also conserved in plant Hbs. The highly conserved F69 of known plant Hbs is a leucine residue in StHb. A motif (N-P-X-L-X-X-H-A-X-X-X-F) is conserved in the both encoded TaHb1 and StHb amino acid sequences. The asparagine residue in the amino terminal of this sequence seems to be conserved in nonsymbiotic Hbs.

An alignment of TaHb2 with 2-on-2 Hbs from barley and *Arabidopsis* is shown in Fig. 2. The alignment reveals a remarkably high homology between the wheat and barley sequences and a significantly lower homology to the *Arabidopsis* 2-on-2 Hb. A phylogenetic analysis of the four different 2-on-2 Hbs demonstrated that barley, wheat and maize group together and different from the *Arabidopsis* GLB3 (data not shown).

The hydropathy profiles of the predicted TaHb1 and StHb proteins were determined by computer analysis using the protscale program (http://www.expasy.ch/cgi-bin/protscale). The hydropathy analysis of TaHb1 and StHb (data not presented) showed that both have a similar profile to other plant Hbs. This suggests that the predicted tertiary structure of TaHb1 and StHb is predominantly formed by α-helices. The results showed that the hydropathy profiles of TaHb1 and StHb are very similar to that of recombinant rice Hb1 indicating that the tertiary structure of TaHb1 and StHb is identical to that of rice Hb1. The crystal structure of a nonsymbiotic rice Hb has recently been reported and the elucidated tertiary structure of the rice Hb1 demonstrated it to be a dimer [21].

In an effort to determine the evolutionary relationships between members of the Hbs, a phylogenetic tree was generated using computer software (MEGALIGN program, Clustal method). A phylogenetic analysis of 13 different plant Hbs has been carried out (Fig. 3). As seen in the evolutionary tree of the amino acid sequences, Hbs separate into four groups reflecting their evolution. The phylogenetic analysis shows that TaHb1 clusters with other monocot Hbs and forms a group distinct from nonsymbiotic dicot Hbs. Three distinct groups contain all dicot Hbs except cotton Hb. The wheat Hb1 clearly belongs to the monocotyledonous clade clustering together with barley and maize Hb and groups individually and different from the rice Hb1 and Hb2. The potato Hb is found together with tomato Hb but, surprisingly, groups with all the monocot nonsymbiotic Hb polypeptide sequences. It is noteworthy that all other dicot Hbs fall into another side of the root and that the cotton Hb seems to belong to a third distinct group of nonsymbiotic Hbs.

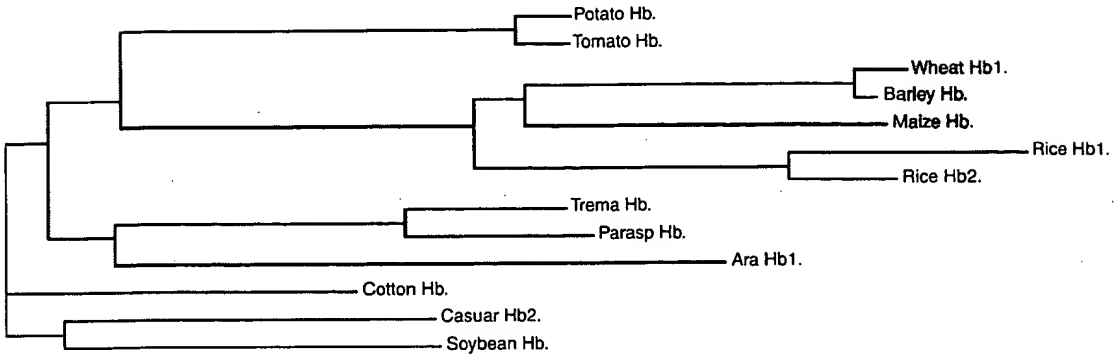

Fig. 3. Phylogenetic tree of TaHb1, StHb and other plant Hbs. The tree was constructed using the using the clustal method of DNASTAR Megalign (DNASTAR, Inc., Madison, WI) based on amino acid similarities of the full sequences. The accession numbers of the sequences used for construction of the phylogenetic tree are listed in the legend to Fig. 2.

To determine the spatial expression of the Hb genes isolated from wheat, *TaHb1* and *TaHb2*, semiquantitative RT-PCR analysis was carried out with total RNA extracted from root, stem, seminal leaf, leaf and flowers of wheat. As illustrated in Fig. 4A by a strong fluorescent 516-bp band, the RT-PCR analysis demonstrated a very high expression level of *TaHb1* in wheat root tissue and a very low level of expression in cotyledons. No TaHb1 transcript could be detected in stems, leaves and flowers of wheat. The RT-PCR analysis also revealed a *TaHb2* transcript of 541 bp in all organs of wheat, the level of expression varying between organs (Fig. 4B). *TaHb2* transcript is more abundant in flower, stems and leaves than in roots.

A constitutive expression was also demonstrated for another 2-on-2 Hb, GLB3 from *A. thaliana* with a slightly increased expression in roots [22]. In leaves, transcript abundance increases with maturity as almost no *TaHb2* transcript is detected in seminal leaf. The gene-specific primers for ryegrass glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were used as internal controls in order to assure that equal amounts of cDNA were used in each individual RT-PCR experiment. A 380-bp band representing the wheat GAPDH mRNA revealed an expression pattern with equally intense bands, indicating that equal amounts of wheat RNA have been used.

The effect of oxygen concentration on the expression of wheat Hbs and GAPDH was investigated by RT-PCR. Twelve-day-old wheat seedlings were incubated under 16 h of controlled oxygen tensions. Marmalade jars were flushed with high purity N_2 for 10 min and the O_2 concentration was adjusted to appropriate concentrations. The oxygen concentrations were measured at $t_0=0$ h and $t=16$ h using a gas chromatograph (Varian Star 3400CX). A flooding experiment was also conducted with wheat seedlings where the entire root system was completely flooded for 16 h. Control wheat plants were grown on wetted filter paper. Total RNA was isolated from root and

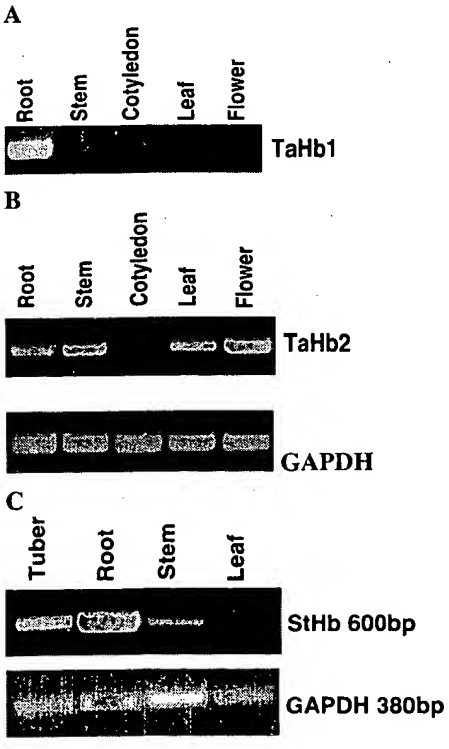

Fig. 4. RT-PCR expression analysis of *TaHb1*, *TaHb2* and *StHb*. The sizes of the *TaHb1*, *TaHb2* and *StHb* and *GAPDH* fragments were 516, 541, 646 and 380 bp, respectively.

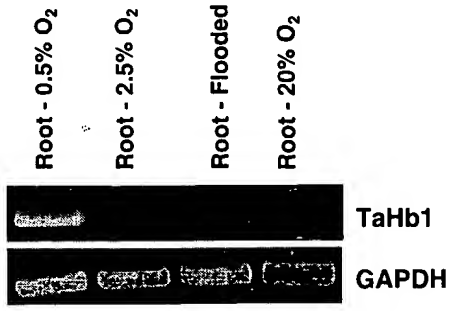

Fig. 5. RT-PCR expression analysis of *TaHb1* in wheat roots exposed to different oxygen concentrations.

leaf of the wheat seedlings as described above and 1 μg was used in an RT-PCR reaction with the same composition as described earlier in this paragraph. The results presented in Fig. 5 clearly demonstrated that the expression of *TaHb1* is induced in wheat roots exposed to anaerobic conditions and that the highest expression level was obtained with an oxygen concentration of 0.5% and a slightly lower expression at 2.5%. *TaHb1* is clearly induced by hypoxia and hence it can be grouped together with class I Hbs. The *GLB1* (earlier nomenclature Ahb1) from *Arabidopsis* and other class I Hbs display similar behaviour [5,17,25]. Surprisingly, the roots of flooding-stressed wheat plants did not show any increased level of *TaHb1* transcript compared to roots from normally grown plants. The lacking response in expression of *TaHb1* in flooded roots can possibly be explained by the experimental design where only the roots were flooded. Either oxygen diffusion in the water covering the roots or oxygen supplied to the roots from the aerial part of the wheat seedling could have created a higher oxygen tension than intended. The expression of both *TaHb2* and *GAPDH* was unaffected of the different anaerobic conditions. The expression of potato Hb, *StHb*, was determined by semiquantitative RT-PCR of total RNA extracted from potato tuber, root, stem and leaf with an expected amplification product of 646 bp As illustrated in Fig. 4C, a high *StHb* expression level was detected in potato root tissue and also in tuber tissue although at a reduced level. A very faint signal representing a low level of potato Hb transcript was observed in stem tissue. No *StHb* transcript could be detected in potato leaf tissue. RT-PCR experiments were also carried out without reverse transcriptase and no fluorescent bands were detected, indicating no contamination with genomic DNA in the RNA preparations (data not shown). The gene-specific primers for ryegrass GAPDH were used as internal controls in order to assure that equal amounts of cDNA were used in each individual RT-PCR experiment. A 380-bp band representing the GAPDH mRNA was represented in root, stem and leaf tissue of potato at the same fluorescence intensity, indicating that equal amounts of cDNA were used.

The cloning of three different cDNAs encoding nonsymbiotic Hbs is reported here. Two Hbs were isolated from wheat and one Hb from potato. The identity of the *Hb* cDNAs was demonstrated by comparison with peptide sequence data from other known plant Hbs. Functional expression of *TaHb1*, *TaHb2* and *StHb* as recombinant proteins is needed for final confirmation of the identity of the Hbs and to verify that the recombinant versions possess the functionality characteristic of Hbs. The data obtained from this work suggest that TaHb1 is important for the development of normal wheat tissues. Under conditions of hypoxia, increased Hb1 synthesis in wheat does not appear to be a general response. Rather, the response may be confined to tissues stressed for energy supply and for oxygen limitations.

Acknowledgements

The author wishes to thank Britta Poulsen for excellent technical assistance, and Drs. Lise Lotte Christensen and Dorothy K. Madsen for the critical reading of the manuscript. Thanks also to Dr. Marianne Madsen, Department of Crop Physiology and Soil Science, Danish Institute of Agricultural Sciences, P.O. Box 50, DK-8830 Tjele Denmark, for providing the potato plants used in this work.

References

[1] S.N. Vinogradov, D.A. Waltz, B. Pohajdak, L. Moens, O.H. Kapp, T. Suzuki, C.N.A. Trotman, Adventitious variability? The amino acid sequences of nonvertebrate globins, Comp. Biochem. Physiol. 106B (1993) 1–26.

[2] R. Weber, S.N. Vinogradov, Nonvertebrate hemoglobins: functions and molecular adaptions, Physiol. Rev. 81 (2001) 569–628.

[3] C.A. Appleby, The origin and functions of haemoglobin in plants, Sci. Prog. 76 (1992) 365–398.

[4] T. Hendriks, I. Scheer, M.C. Quillet, B. Randoux, B. Delbreil, J. Vasseur, J.L. Hilbert, A nonsymbiotic hemoglobin gene is expressed during somatic embryogenesis in *Cichorium*, Biochim. Biophys. Acta 1443 (1998) 193–197.

[5] E.R. Taylor, X.Z. Nie, A.W. MacGregor, R.D. Hill, A cereal haemoglobin gene is expressed in seed and root tissues under anaerobic conditions, Plant Mol. Biol. 24 (1994) 853–862.

[6] P.W. Hunt, R.A. Watts, B. Trevaskis, D.J. Llewelyn, J. Burnell, W.J. Peacock, E.S. Dennis, *Marchantia polymorpha* non-vascular plant hemoglobin (GLB0) gene, partial cds. GenBank AY026341 (2001).

[7] R. Arredondo-Peter, M. Ramirez, G. Sarath, R.V. Klucas, Sequence analysis of an ancient hemoglobin cDNA isolated from the moss *Physcomitrella patens* (accession no. AF218049), Plant Physiol. 122 (2000) 1457.

[8] T. Sasaki, J. Song, Y. Koga-Ban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. Hori, M. Miya, E. Murayama-Kayano, T. Takiguchi, A. Takasuga, T. Niki, K. Ishimaru, H. Ikeda, Y. Yamomato, T. Mukai, I. Ohta, N. Miyadera, I. Havukkala, Y. Minobe, Toward catalogueing all rice genes: large scale sequencing of randomly chosen rice cDNAs from a callus cDNA library, Plant J. 6 (1994) 615–624.

[9] P.A. Guy, R.D. Hill, *Zea mays* hemoglobin mRNA, complete cds. Genbank AF236080 (2000).

[10] E. Aréchaga-Ocampo, J. Saenz-Rivera, G. Sarath, R.V. Klucas, R. Arredondo-Peter, Cloning and expression analysis of hemoglobin genes from maize (*Zea mays* ssp. *mays*) and teosinte (*Zea mays* ssp. *parviglumis*), Biochim. Biophys. Acta 1522 (2001) 1–8.

[11] C.R. Andersson, E.O. Jensen, D.J. Llewellyn, E.S. Dennis, W.J. Peacock, A new hemoglobin gene from soybean: a role for hemoglobin in all plants, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 5682–5687.

[12] B. Trevaskis, R.A. Watts, C.R. Andersson, D.J. Llewellyn, M.S. Hargrove, J.S. Olson, E.S. Dennis, W.J. Peacock, Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 12230–12234.

[13] R.D. Hill, What are hemoglobins doing in plants? Can. J. Bot. 76 (1998) 707–712.

[14] S.M.G. Duff, P.A. Guy, X. Nie, D.C. Dumin, R.D. Hill, Haemoglobin expression in germinating barley, Seed Sci. Res. 8 (1998) 431–436.

[15] A.W. Sowa, S.M.G. Duff, P.A. Guy, R.D. Hill, Altering hemoglobin levels changes energy status in maize cells under hypoxia, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 10317–10321.

[16] V. Lira-Ruan, G. Sarath, R.V. Klucas, R. Arredondo-Peter, Synthesis of hemoglobins in rice (*Oryza sativa* var. Jackson) plants growing in normal and stress conditions, Plant Sci. 161 (2001) 279–287.

[17] X. Nie, R.D. Hill, X.Z. Nie, Mitochondrial respiration and haemoglobin gene expression in barley aleurone tissue, Plant Physiol. 114 (1997) 835–840.
[18] C. Seregelyes, L. Mustardy, F. Ayaydin, L. Sass, L. Kovacs, G. Endre, N. Lukacs, I. Kovacs, I. Vass, G.B. Kiss, G.V. Horvath, D. Dudits, Nuclear localization of a hypoxia-inducible novel non-symbiotic hemoglobin in cultured alfalfa cells, FEBS Lett. 482 (2000) 125–130.
[19] E.J.H. Ross, V. Lira-Ruan, R. Arredondo-Peter, R.V. Klucas, C. Sarath, Recent insights into plant hemoglobins, Rev. Plant Biochem. Biotechnol. 1 (2002) 173–189.
[20] C. Dordas, J. Rivoal, R.D. Hill, Plant haemoglobins, nitric oxide and hypoxic stress, Ann. Bot. 91 (2003) 173–178.
[21] M.S. Hargrove, E.A. Brucker, B. Stec, G. Sarath, R. Arredondo-Peter, R.V. Klucas, J.S. Olson, G.N. Philips Jr., Crystal structure of a non-symbiotic plant hemoglobin, Structure 8 (2000) 1005–1014.
[22] R.A. Watts, P.W. Hunt, A.N. Hvitved, M.S. Hargrove, W.J. Peacock, E.S. Dennis, A hemoglobin from plants homologous to truncated hemoglobins of microorganisms, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10119–10124.
[23] A. Pesce, M. Couture, S. Dewilde, M. Guertin, K. Yamauchi, P. Ascenzi, L. Moens, M. Bolognesi, A novel two-over-two alpha-helical sandwich fold is characteristic of the truncated hemoglobin family, EMBO J. 19 (2000) 2424–2434.
[24] W.H. Campbell, G. Gowri, Codon usage in higher plants, green algae and cyanobacteria, Plant Physiol. 92 (1990) 1–11.
[25] P.W. Hunt, R.A. Watts, B. Trevaskis, D.J. Llewelyn, J. Burnell, E.S. Dennis, W.J. Peacock, Expression and evolution of functionally distinct haemoglobin genes in plants, Plant Mol. Biol. 47 (2001) 677–692.

doi: 10.1046/j.1420-9101.2004.00658.x

Molecular evolution of plant haemoglobin: two haemoglobin genes in nymphaeaceae *Euryale ferox*

E. GULDNER, E. DESMARAIS, N. GALTIER & B. GODELLE

CNRS UMR 5000 – 'Génome, Populations, Interactions', Université Montpellier, Montpellier, France

Keywords:

duplication;
nonsynonymous substitution rate;
Nymphaeaceae;
plant haemoglobin;
symbiosis.

Abstract

We isolated and sequenced two haemoglobin genes from the early-branching angiosperm *Euryale ferox* (Nymphaeaceae). The two genes belong to the two known classes of plant haemoglobin. Their existence in Nymphaeaceae supports the theory that class 1 haemoglobin was ancestrally present in all angiosperms, and is evidence for class 2 haemoglobin being widely distributed. These sequences allowed us to unambiguously root the angiosperm haemoglobin phylogeny, and to corroborate the hypothesis that the class 1/class 2 duplication event occurred before the divergence between monocots and eudicots. We addressed the molecular evolution of plant haemoglobin by comparing the synonymous and nonsynonymous substitution rates in various groups of genes. Class 2 haemoglobin genes of legumes (functionally involved in a symbiosis with nitrogen-fixing bacteria) show a higher nonsynonymous substitution rate than class 1 (nonsymbiotic) haemoglobin genes. This suggests that a change in the selective forces applying to plant haemoglobins has occurred during the evolutionary history of this gene family, potentially in relation with the evolution of symbiosis.

Introduction

Two types of haemoglobin, called class 1 and class 2 haemoglobin, occur in vascular plants. They are distinct from each other through their function and expression pattern (Jacobsen-Lyon *et al.*, 1995; Trevaskis *et al.*, 1997; Hunt *et al.*, 2001). Class 1 haemoglobins have been found in both monocots and eudicots and show a conserved expression pattern (Hunt *et al.*, 2001). They are expressed in germinating seeds [barley (Duff *et al.*, 1998)], in roots of mature plants (*Parasponia, Trema* and *Casuarina* GLB1 (Bogusz *et al.*, 1988; Bogusz *et al.*, 1990; Jacobsen-Lyon *et al.*, 1995; Andersson *et al.*, 1997; Franche *et al.*, 1998) and rosette leaves subjected to hypoxic conditions, and in response to sucrose [*Arabidopsis* (Trevaskis *et al.*, 1997)]. Class 2 haemoglobins were first described in plants achieving symbiosis with nitrogen-fixing bacteria (symbiotic haemoglobins). In most legume species (symbiosis with *Rhizobium* and *Bradyrhizobium*) and in several species from Fagales, Rosales and Cucurbitales (symbiosis with *Frankia*), class 2 haemoglobins are expressed at high concentration in nodules and facilitate oxygen diffusion to the nitrogen-fixing symbionts (Appleby, 1992). Class 2 haemoglobins were also found in several nonsymbiotic eudicots (nonsymbiotic haemoglobins), where their function remains unclear. In these species, they are expressed in several tissues (Szabados *et al.*, 1990; Strozycki *et al.*, 2000), and might participate in embryogenesis and seed maturation [*Cichorium* (Hendriks *et al.*, 1998), *Arabidopsis* (Hunt *et al.*, 2001)]. At this moment no class 2 haemoglobin gene has been found in any monocot.

Expression patterns, therefore, suggest that the evolution of haemoglobin may have played a role in the evolution of symbiosis between angiosperms and nitrogen-fixing bacteria, stimulating research about the molecular phylogeny and evolution of this gene family. Hypotheses about plant haemoglobin evolution have gradually progressed thanks to the characterization of new sequences (Andersson *et al.*, 1996; Trevaskis *et al.*, 1997; Arredondo-Peter *et al.*, 1998; Strozycki *et al.*, 2000). We are a long way from the first postulate of a horizontal transfer of the haemoglobin gene from the

Correspondence: Guldner Emilie, CNRS UMR 5000 – 'Génome, Populations, Interactions', Université Montpellier 2 – CC 063, Place E. Bataillon, 34095 Montpellier, France. Fax: +33 467 14 45 54; e-mail: guldner@univ-montp2.fr

animal kingdom to a progenitor of the dicotyledonous angiosperms. Molecular phylogeny analyses support the grouping of class 1 vs. class 2 haemoglobins (Andersson *et al.*, 1996; Trevaskis *et al.*, 1997; Arredondo-Peter *et al.*, 1998; Strozycki *et al.*, 2000), the two clusters being separated by a long internal branch. The recent identification of haemoglobin genes in nonvascular plants (liverwort *Marchantia polymorpha* (Hunt *et al.*, 2001) and mosses *Physcomitrella patens* (Arredondo-Peter *et al.*, 2000) and *Ceratodon purpureus* (accession number AF309562) provides a way to root the angiosperm haemoglobin phylogeny, potentially casting light on the evolutionary history of this gene family. According to Hunt *et al.* (2001), the root of the tree takes place between the class 1 and class 2 clusters (Fig. 1a), suggesting that class 1 and class 2 haemoglobins arose through an ancient duplication event that occurred before the divergence between monocots and eudicots. This hypothesis requires that class 2 genes have been lost in the monocot lineage.

Nonvascular plants, however, are quite divergent from angiosperms (Soltis *et al.*, 1999), casting doubts on the location of the root of the haemoglobin tree. The so-called 'long branch attraction' artefact (Felsenstein, 1978) may be responsible for the moss sequences being connected to the long branch that separates class 1 from class 2 haemoglobins, rather than to any other branch in the un-rooted tree. An alternative scenario involving a eudicot-specific duplication event, and paraphyletic class 1 haemoglobins, would be more parsimonious with respect to gene gain/loss (Fig. 1b).

Also, the published haemoglobin trees suggest that the rate of amino-acid substitution varies between lineages (Strozycki & Legocki, 1995; Strozycki *et al.*, 2000). The divergence between class 2 haemoglobins from various legume species, for example, appears comparable with the divergence between monocot and eudicot class 1 genes despite a more restricted taxonomic range, suggesting a possible acceleration of evolutionary rate in the class 2 lineage. This acceleration, if confirmed, would be an interesting evolutionary feature of the class 2 symbiotic genes. It might, however, contribute to possible phylogenetic artefact.

These arguments emphasize the need for a solid rooting of the haemoglobin tree. This would help in clarifying the phylogenetic history of the gene family in symbiotic and nonsymbiotic plants. Furthermore, the measurement of rate variation between lineages being highly dependent on the assumed phylogeny and rooting, this would also allow assessing the extent of evolutionary rate variation between lineages, possibly connecting features of the molecular evolutionary process in haemoglobin to the evolution of symbiosis. In this study, we made use of Nymphaeaceae *Euryale ferox*, a deeply-branching lineage of the angiosperm tree, to root the haemoglobin tree, and reconsider the molecular evolution of plant haemoglobins.

Material and methods

Young leaves from the Nymphaeaceae *E. ferox* were collected in the botanical garden of Montpellier (France). DNA extraction was performed with the DNeasy Plant Kit (QIAGEN, Courtaboeuf, France).

Degenerate primers for polymerase chain reaction (PCR) were defined from aligned plant haemoglobin sequences available in GenBank, and used to amplify each haemoglobin gene in two overlapping regions. Class 1 primers are: TTCAGYGARGAGCARGARGC (exon 1 forward)/GTGGWCTCCCTCACSGTRAC (exon3 reverse), and GACTCCGACCTGCCACTTGAAC (exon 2 forward)/GCRGCBACCARYTGRTCRTARGC (exon 4 reverse). Class 2 primers are: TTCAGYGARGAGCARGARGC (exon 1 forward)/GTRGCRCCYAGYCTYTTCAG (exon 3 reverse), and AACCCSMARCTCARGYCYCAT (exon 2 forward)/GCMRCCAACTKRTCRTARGC (exon 4 reverse). The PCR reactions included *Taq* DNA polymerase buffer (Promega)/1.5 mM MgCl2/200 μM each dNTP/

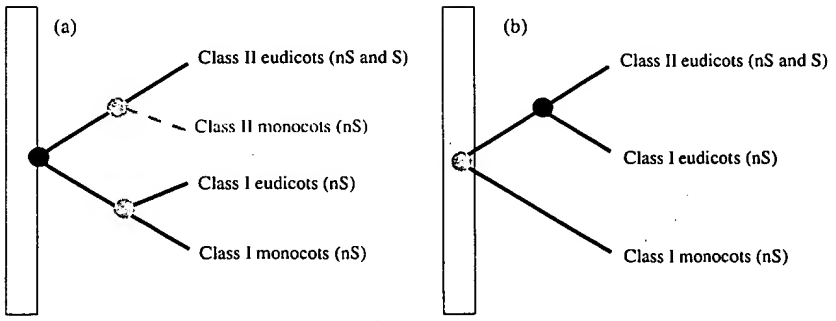

Fig. 1 Evolutionary history of plant haemoglobin genes: competing hypothesis. (a) A gene duplication gave rise to class 1 and class 2 haemoglobins before the divergence between monocots and eudicots. Class 2 haemoglobins have been lost (or are still undetected) in cereals. (b) A gene duplication occurred in the eudicot lineage – black dot: gene duplication; gray dot: monocots/eudicots divergence; S: symbiotic gene; nS: nonsymbiotic gene.

variable quantities of each primer, depending on their degeneracy/genomic DNA/5 units *Taq* DNA polymerase (Promega) in a total volume of 50 μL. The reaction consists of a first step at 94 °C for 2 min, followed by 30 cycles with a denaturation step at 91 °C for 30 s, an annealing step at variable temperature (from 55 °C to 62 °C) for 20 s and an extension step at 72 °C for 30 s.

The amplified fragments, purified with the GenElute PCR DNA purification kit (Sigma-Aldrich, Saint Quentin Fallavier, France), were cloned with the pGEM-T Easy Vector System (Promega). For each amplified fragment, four to six clones were sequenced using the ThermoSequenase sequencing kit with dye primer (Amersham Pharmacia Biosciences, Orsay, France) following the provider's instructions. The sequences were analysed with the ALFwin Sequence Analyser (Amersham Pharmacia Biosciences).

Phylogenetic analyses were performed on 52 plant haemoglobin amino-acid sequences (143 sites, 132 sites with gaps removing, 121 variable sites, 110 informative sites). We used methods of maximum likelihood [using the JTT model (Jones *et al.*, 1992)] with the program NJML+ (Ota & Li, 2001), and Neighbour-Joining (NJ, Saitou & Nei, 1987) with the program PHYLO_WIN (Galtier *et al.*, 1996), with 1000 bootstrap replicates for each analysis. Phylogenetic analysis based on Bayesian inference was also performed with *MrBayes* program (Huelsenbeck & Ronquist, 2001). We used the JTT model assuming a discretized gamma distribution with four classes. 120 000 Monte Carlo Markov chain (MCMC) steps were performed using four Markov chains running at different temperatures. The 20 000 first steps were discarded (burnin). Trees were sampled every 100 steps, and a consensus tree (out of 1000 sampled trees) was constructed. Clade posterior probabilities allow evaluation of the robustness of each node.

Pairwise distances between protein sequences were calculated using the PAM correction for multiple substitutions (Dayhoff *et al.* 1978), as implemented in program PROTDIST from package PHYLIP (Felsenstein 1989). These distances, used to build the NJ tree, were also used to examine sequence similarities between the newly isolated sequences of *E. ferox* and others plant haemoglobin genes.

The heterogeneity of evolution rate between lineages was measured using a generalization of the relative-rate test (Wu & Li, 1985) to more than three species (Robinson-Rechavi *et al.*, 1998; Robinson-Rechavi & Huchon, 2000). The synonymous (*Ks*) and nonsynonymous (*Ka*) evolutionary distances were estimated for each sequence pair within the group of eudicot class 1 genes (11 genes from eight different species, 55 pairs), the group of monocot class 1 genes (four genes from three different species, six pairs), and the group of class 2 (symbiotic) genes of legumes (27 genes from 11 different species, 351 pairs). We used the method of Yang & Nielsen (2000) as implemented in program yn00 of PAML (Yang, 1997).

Results and discussion

Two haemoglobin sequences in Nymphaeaceae *E. ferox*

All the fragments produced by degenerate amplifications on the *E. ferox* genomic DNA were cloned and sequenced. Two sequences (called *Hbn1* and *Hbn2*) showed significant similarity to known haemoglobin sequences. *Hbn2* shows three introns and four exons in positions identical to those of other plant haemoglobins (Jensen *et al.*, 1981; Landsmann *et al.*, 1986). Concerning *Hbn1*, amino-acid sequence similarity to known haemoglobin sequences was recovered when a nonstandard intron–exon map was used: exon 3 appears 42 nucleotides longer (and exon 4, 42 nucleotides shorter) than in other plant haemoglobin genes. PAM distances on amino acid sequences were estimated between each of these two new genes and all others known sequences. *Hbn1* is much closer to the class 1 haemoglobins (average PAM distance: 0.41 replacements per site) than to class 2 nonsymbiotic (0.86 rep. per site) and symbiotic (1.19 rep. per site) haemoglobins. In contrast, *Hbn2* is more similar to class 2 nonsymbiotic genes (0.76 rep. per site) than to class 1 (0.90 rep. per site) and symbiotic (1.05 rep. per site) haemoglobins. These results show that at least two haemoglobin genes actually exist in *E. ferox*, presumably belonging to class 1 (*Hbn1*) and class 2 (*Hbn2*), respectively.

Figure 2 presents the phylogenetic tree of 52 plant haemoglobin genes, constructed from 132 amino-acid sites with phylogenetic analysis based on Bayesian inference. Both groups of class 1 and class 2 genes are supported by high clade posterior probabilities (0.97 and 1.00 respectively). Phylogenetic reconstructions obtained using the NJ and maximum of likelihood methods are congruent with the Bayesian tree in that they present two strongly supported clades composed of class 1 genes (including *Hbn1*), and class 2 genes (including *Hbn2*) respectively. Note that both the class 1 and class 2 genes of *E. ferox* branch out as the earliest diverging lineages, in agreement with recent angiosperm molecular phylogenies (Soltis *et al.*, 1999; Soltis *et al.*, 2000). These results unambiguously confirm the previous rooting of the haemoglobin tree (Hunt *et al.*, 2001), i.e. the gene duplication that gave rise to class 1 and class 2 haemoglobins is older than the split between monocots and eudicots. Thus, class 2 haemoglobins must have been lost (or are still undetected) in cereals, and more research is needed to find out whether the lack of class 2 haemoglobin is restricted to cereals, or general to all monocots.

The unambiguous rooting of the tree gives an opportunity to compare molecular evolutionary rates between haemoglobin lineages.

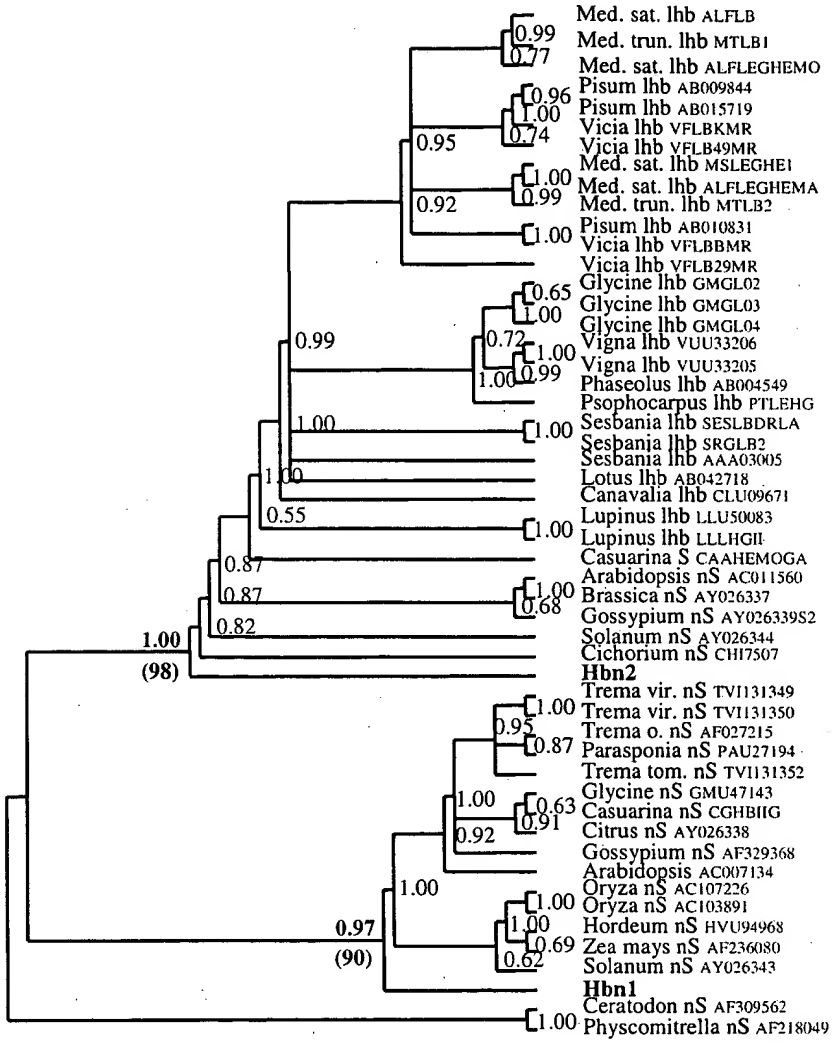

Fig. 2 Bayesian tree: 52 species, 132 amino acid sites. Clade posterior probabilities are given for each node. Neighbour-Joining bootstrap supports (1000 replicates) are indicated within parenthesis. lhb: symbiotic haemoglobin of legumes; S: nonlegume symbiotic gene; nS: nonsymbiotic gene.

Evolutionary rates of plant haemoglobins

As long as it is assumed that the right topology has been inferred, the estimation of the branch lengths of a phylogeny may provide useful information about molecular evolutionary processes. In the case of the plant haemoglobin phylogeny, several nodes remain unresolved within each group of genes (Fig. 2). However, at this time we have sufficiently reliable knowledge of angiosperm phylogeny (Soltis *et al.*, 1999; Soltis *et al.*, 2000), which can help to resolve some of these ambiguities. We thus modified the haemoglobin gene phylogeny to make it in agreement with the canonical angiosperm phylogeny, grouping together genes belonging to the same family or order. No change was made within the legume class 2 group, in which many recent duplication events have occurred, confusing species phylogeny. Branch lengths of the haemoglobin gene tree were then estimated by fitting PAM distances to the modified topology (least square fit). This 'true' haemoglobin phylogeny (Fig. 3) shows longer branches for the class 2 sequences, and especially for symbiotic genes. The level of significance of this trend was assessed using a generalization of the relative-rate test to more than three species (Robinson-Rechavi *et al.*, 1998). The difference of amino-acid substitution rates between class 1 and class 2 genes, using moss sequences as out-group, was found to be highly significant ($P < 0.001$).

The acceleration observed for class 2 genes might have been caused by an increase of mutation rate (e.g. if located in a rapidly evolving genomic region). Alternatively, this pattern might reflect a change in the selective forces applying to the haemoglobin gene. To distinguish between these two hypotheses, we contrasted the synonymous (*Ks*) and nonsynonymous (*Ka*) evolutionary rates for each sequence pair within two groups of class 1 genes (eudicots and monocots respectively) on one hand, and the group of symbiotic class 2 genes of legumes on

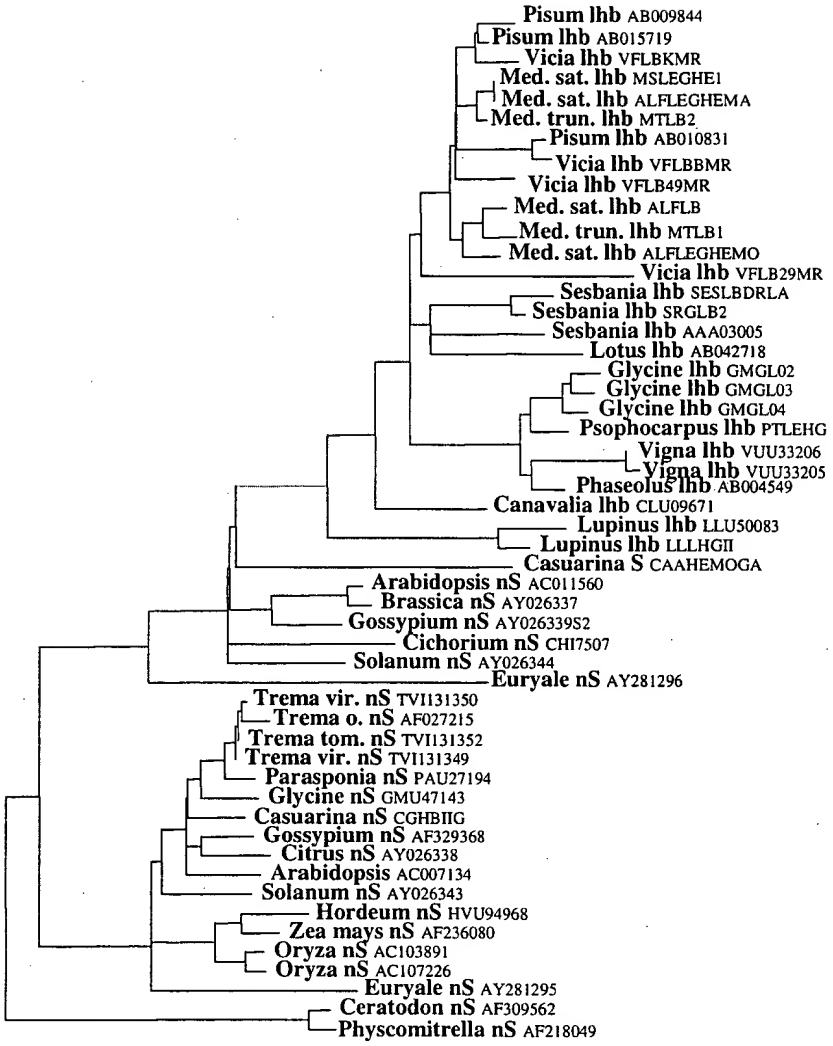

Fig. 3 Haemoglobin gene phylogeny: 52 species, 132 amino acid sites. Branch lengths were estimated by fitting PAM distances between amino-acid sequences to a modified haemoglobin tree topology (see text).

the other hand. Under the hypothesis of a variable mutation rate but a constant selective regime, an equal *Ks/Ka* ratio would be expected in all groups. Figure 4 leads to rejection of this hypothesis: the *Ks/Ka* ratio is much lower in the class 2 symbiotic group (average: 4.6) than in class 1 groups (average: 9.3 for eudicots and 12.5 for monocots). This indicates that the selective forces applying (or having applied) to symbiotic class 2 haemoglobins are distinct from those applying to class 1 haemoglobins.

These within-group comparisons do not reflect the evolutionary process that accompanied the functional divergence of the two genes after the duplication, but rather what occurred after their current function was acquired. Again, two main hypotheses might be invoked to explain the observed difference in *Ks/Ka* ratio between the two lineages. First, a lower *Ks/Ka* in class 2 genes might result from the relaxation of functional constraints, i.e. a decrease of the strength of purifying selection. Amino acids might be freer to vary in class 2 than in class 1 genes, for a yet undetermined reason. This is the simplest explanation, consistent with the neutralist theory, and should be considered as the null hypothesis. Alternatively, recurrent adaptation might also explain an increase of nonsynonymous substitution rate in class 2 symbiotic genes. Advantageous (nonsynonymous) mutations have a higher fixation probability than neutral mutations, and contribute to decreasing the *Ks/Ka* ratio.

The reason why the adaptive hypothesis could be considered as a reasonable alternative to the neutralist explanation in this case is that the legume haemoglobins are functionally involved in the symbiosis between legumes and Rhizobium. Interactions between species potentially generate the need for a recurrent adaptation (Red-Queen-like evolution), one species having to adapt to the innovations found by the other one. Nitrogen

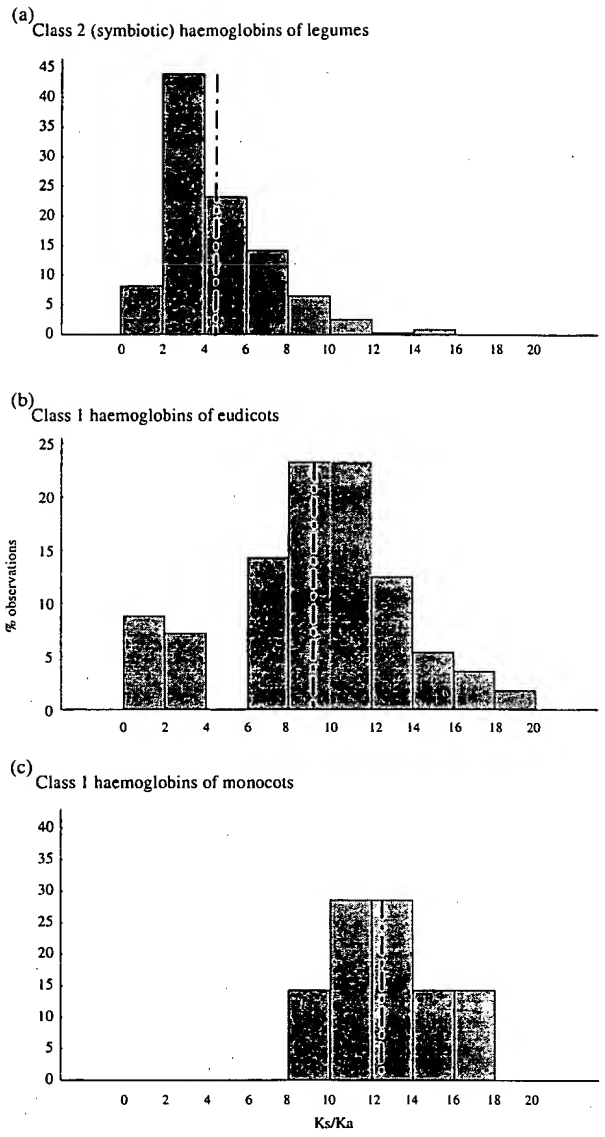

Fig. 4 Distribution of pairwise *Ks/Ka* ratios for legume symbiotic genes (a), eudicots class 1 genes (b) and monocots class 1 genes (c). Within each group, synonymous and nonsynonymous distances were estimated for all sequence pairs. The vertical lines indicate the average values.

fixation by bacteria, for instance, involves the nitrogenase enzyme, which requires substantial amounts of energy (i.e. ATP) produced by bacteroid respiration. However, the oxygen necessary for respiration readily inhibits the activity of nitrogenase. Haemoglobin thus appears to be a key factor in the process of symbiotic exchange by facilitating oxygen diffusion to the nitrogen-fixing symbionts (Appleby, 1992). From an evolutionary perspective, increasing or decreasing the affinity of haemoglobin to oxygen might therefore be a way for legumes to control the bacterial growth and amount of symbiotic exchange or to adapt to a new strain of bacteria (Bever & Simms, 2000; Denison, 2000; West *et al.*, 2002).

Adaptation at the sequence level can definitely be invoked when *Ka* is higher than *Ks*, a pattern incompatible with neutral evolution or purifying selection. This is not what we observed for haemoglobin. Not observing *Ka/Ks*, however, does not imply that adaptation has not occurred. Indeed, this adaptive evolution might have involved only a (small) fraction of sites, and/or have occurred in specific branches of the tree. *Ks/Ka* measurements averaged over sites and over lineages make the detection of episodes of adaptive evolution difficult, and additional analyses involving models allowing for variable *Ks/Ka* ratio among sites/branches (Goldman & Yang, 1994; Yang, 1997; Yang, 2000) are required for distinguishing between the two hypotheses.

References

Andersson, C.R., Jensen, E.O., Llewellyn, D.J., Dennis, E.S. & Peacock, W.J. 1996. A new haemoglobin gene from soybean : a role for haemoglobin in all plants. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 5682–5687.

Andersson, C.R., Llewellyn, D.J., Peacock, W.J. & Dennis, E.S. 1997. Cell-specific expression of the promoters of two nonlegume haemoglobin genes in a transgenic legume, *Lotus corniculatus*. *Plant Physiol.* **113**: 45–57.

Appleby, C. 1992. The origin and functions of haemoglobin in plants. *Sci. Progress* **76**: 365–398.

Arredondo-Peter, R., Hargrove, M.S., Moran, J.F., Sarath, G. & Klucas, R.V. 1998. Plant haemoglobins. *Plant Physiol.* **118** 1121–1125.

Arredondo-Peter, R., Ramirez, M., Sarath, G. & Klucas, R.V. 2000. Sequence analysis of an ancient haemoglobin cDNA isolated from the moss *Physcomitrella patens*. *Plant Physiol.* **122**: 1457.

Bever, J.D. & Simms, E.L. 2000. Evolution of nitrogen fixation in spatially structured populations of Rhizobium. *Heredity* **85**: 366–372.

Bogusz, D., Appleby, C., Landsmann, J., Dennis, E.S., Trinick, M.J. & Peacock, W.J. 1988. Functioning haemoglobin genes in non-nodulating plants. *Nature* **331**: 178–180.

Bogusz, D., Llewellyn, D.J., Craig, S., Dennis, E.S., Appleby, C. & Peacock, W.J. 1990. Nonlegume haemoglobin genes retain organ-specific expression in heterologous transgenic plants. *The Plant Cell* **2**: 633–641.

Dayhoff, M.O., Schwartz, R.M. & Orcutt, B.C. 1978. A model of evolutionary change in proteins; Matrices for detecting distant relationships. In: *Atlas of protein sequence and structure*, pp. 345–358. National Biomedical Research Foundation, Washington, DC.

Denison, R.F. 2000. Legume sanctions and the evolution of symbiotic cooperation by Rhizobia. *Am. Nat.* **156**: 567–576.

Duff, S.M.G., Guy, P.A., Nie, X., Durnin, D.C., Hill, R.D. & Nie, W.Z. 1998. Haemoglobin expression in germinating barley. *Seed Sci. Res.* **8**: 431–436.

Felsenstein, J. 1978. The number of evolutionary trees. *Syst. Zool.* **27**: 27–33.

Felsenstein, J. 1989. PHYLIP – Phylogeny Inference Package (version 3.2). *Cladistics* **5**: 164–166.

Franche, C., Diouf, D., Laplaze L, Auguy, F., Frutz, T., Rio, M., Duhoux, E. & Bogusz, D. 1998. Soybean (*lbc3*), *Parasponia*, and *Trema* haemoglobin gene promoters retain symbiotic specificity in transgenic Casuarinaceae : implications for haemoglobin gene evolution and root nodule symbioses. *Mol. Plant-Microbe Interact* **11**: 887–894.

Galtier, N., Gouy, M. & Gautier, C. 1996. SEAVIEW and PHYLO_WIN : two graphic tools sequence alignment and molecular phylogeny. *CABIOS* **12**: 543–548.

Goldman, N. & Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Biol. Mol. Evol.* **11**: 725–736.

Hendriks, T., Scheer, I., Quillet, M.-C., Randoux, B., Delbreil, B., Vasseur, J. & Hilbert, J.-L. 1998. A nonsymbiotic haemoglobin gene is expressed during somatic embryogenesis in *Cichorium*. *Biochemica et Biophysica Acta* **1143**: 193–197.

Huelsenbeck, J. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Hunt, R.A., Watts, R.A., Trevaskis, B., Llewellyn, D.J., Burnell, J., Dennis, E.S. & Peacock, W.J. 2001. Expression and evolution of functionally distinct haemoglobin genes in plants. *Plant Mol. Biol.* **47**: 677–692.

Jacobsen-Lyon, K., Jensen, E.O., Jorgensen, J.-E., Marcker, K.A., Peacock, W.J. & Dennis, E.S. 1995. Symbiotic and nonsymbiotic haemoglobin genes of *Casuarina glauca*. *The Plant Cell* **7**: 213–223.

Jensen, E.O., Paludan, K., Hyldig-Nielsen, J.J., Jorgensen, P. & Marcker, K.A. 1981. The structure of a chromosomal leghaemoglobin gene from soybean. *Nature* **291**: 677–679.

Jones, D.T., Taylor, W.R. & Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.

Landsmann, J., Dennis, E.S., Higgins, T.J.V., Appleby, C., Kortt, A. & Peacock, W.J. 1986. Common evolutionary origin of legume and non-legume plant haemoglobins. *Nature* **324**: 166–168.

Ota, S. & Li, W. 2001. NJML+ : an extension of the NJML method to handle protein sequence data and computer software implementation. *Mol. Biol. Evol.* **18**: 1983–1992.

Robinson-Rechavi, M. & Huchon, D. 2000. RRTree : relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**: 296–297.

Robinson-Rechavi, M., Gouy, M., Gautier, C. & Mouchiroud, D. 1998. Sensitivity of the relative-rate test to taxonomic sampling. *Mol. Biol. Evol.* **15**: 1091–1098.

Saitou, N. & Nei, M. 1987. The neighbor-joining method : a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.

Soltis, P.S., Soltis, D.E. & Chase, M.W. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**: 402–403.

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., Axtel, M., Swensen, S.M., Prince, L., Kress, W.J., Nixon, K.C. & Farris, J. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atp*B sequences. *Bot. J. Linn. Soc.* **133**, 381–461.

Strozycki, P.M. & Legocki, A.B. 1995. Leghaemoglobins from an evolutionarily old legume, *Lupinus luteus*. *Plant Science* **110**: 83–93.

Strozycki, P.M., Karlowski, W.M., Dessaux, Y., Petit, A. & Legocki, A.B. 2000. Lupine *leghaemoglobin I* : expression in transgenic *Lotus* and tobacco tissues. *Mol. Gen. Genet.* **263**: 173–182.

Szabados, L., Ratet, P., Grunenberg, B. & Bruijn, F.G. 1990. Functional analysis of the *Sesbania rostrata* leghaemoglobin *glb3* gene 5'-upstream region in transgenic *Lotus corniculatus* and *Nicotiana tabacum* plants. *Plant Cell* **2**: 973–986.

Trevaskis, B., Watts, R.A., Andersson, C.R., Llewellyn, D.J., Hargrove, M.S., Olson, J.S., Dennis, E.S. & Peacock, W.J. 1997. Two haemoglobin genes in Arabidopsis thaliana : the evolutionary origins of leghaemoglobins. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 12230–12234.

West, S.A., Kiers, E.T., Simms, E.L. & Denison, R.F. 2002. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. Natl. Acad. Sci. U. S. A.* **269**: 685–694.

Wu, C. & Li, W.H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. U. S. A.* **82**: 1741–1745.

Yang, Z. 1997. PAML : a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**: 555–556.

Yang, Z. (2000). *Phylogenetic Analysis by Maximum Likelihood (PAML)*, Version 3.0. University College London, London, England (http://abacus.gene.ucl.ac.uk/software/paml.html)

Yang, Z. & Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**: 32–43.

Received 23 May 2003; revised 27 September 2003; accepted 27 September 2003